*UDC 811.111'37'42*
*DOI https://doi.org/10.28925/2412-2491.2026.262*

# COCA AS A REPRESENTATIVE CORPUS FOR THE STUDY OF BASIC EMOTIONS IN CONTEMPORARY ENGLISH

**Nataliia Bober**
Borys Grinchenko Kyiv Metropolitan University
ORCID ID: https://orcid.org/0000-0002-9639-0562
n.bober@kubg.edu.ua



*The article explores six nominal labels for basic emotions in contemporary English, based on the Corpus of Contemporary American English (COCA): anger, fear, sadness, happiness, disgust, and surprise. The selected nouns are not interpreted as direct evidence of inner affective experience. They are analysed as lexical units whose corpus profile is formed by recurrent co-occurrence, grammatical relations, collocational environment, clusters, and concordance contexts. The study combines procedures of corpus linguistics with lexical-semantic interpretation and linguistic research on emotion. The theoretical basis includes work on corpus representativeness, lexical semantics, the theory of basic emotions, and the distinction between emotion-label words and emotion-laden words. COCA was selected because its register-balanced and chronologically stratified design makes it suitable for analysing contemporary American English usage. The material was obtained using the COCA tools Word, Topics, Collocates, Clusters, Texts/Virtual Corpora, and Concordance Lines. The frequency distribution in the six-item sample is as follows: fear - 103,493; surprise - 60,286; anger - 36,471; happiness - 22,057; sadness - 8,851; disgust - 5,436. The analysis prioritises not frequency lists in isolation, but recurrent lexico-grammatical patterns. Fear occurs mainly with nominal and clausal complementation; anger is characterized by prepositional directionality and verbs of expression or regulation; sadness is profiled through intensifiers and perceptible signs; happiness regularly enters axiological contexts of well-being, aspiration, and life purpose; disgust combines bodily aversion with moral evaluation; and surprise functions both as a nominal label for an unexpected event and as a marker of its expectedness or unexpectedness. The findings indicate that COCA makes it possible to describe nominal emotion labels not merely as dictionary entries, but as contextually profiled units of the language system.*

***Key words:*** *COCA, basic emotions, nominal emotion labels, corpus linguistics, collocation, valency*

***Бобер Н.М. Корпус COCA як репрезентативна база дослідження базових емоцій сучасної англійської мови.*** *У статті проаналізовано шість іменникових номінацій базових емоцій у сучасній англійській мові на матеріалі Corpus of Contemporary American English (COCA): anger, fear, sadness, happiness, disgust і surprise. Відібрані іменники не ототожнюються з безпосереднім виявом внутрішнього афективного досвіду мовця. Їх розглянуто як лексичні одиниці, корпусний профіль яких формується*

*через повторювану сполучуваність, граматичні зв'язки, колокаційне оточення, кластери та конкордансні контексти. Методологічно дослідження поєднує корпусно-лінгвістичні процедури з лексико-семантичним аналізом і лінгвістикою емоцій. Теоретичну основу становлять праці з корпусної репрезентативності, лексичної семантики, теорії базових емоцій, а також дослідження, у яких розмежовано лексеми прямої емоційної номінації (emotion-label words) і лексеми з емоційно-оцінним навантаженням (emotion-laden words). Вибір COCA зумовлений його жанровою збалансованістю, хронологічною стратифікацією та придатністю для аналізу сучасного американського слововживання. Матеріал дослідження отримано за допомогою інструментів COCA Word, Topics, Collocates, Clusters, Texts / Virtual Corpora та Concordance Lines. Частотний розподіл у шестикомпонентній вибірці є таким: fear – 103 493; surprise – 60 286; anger – 36 471; happiness – 22 057; sadness – 8 851; disgust – 5 436. Аналітичний акцент зроблено не на ізольованих частотних списках, а на повторюваних лексико-граматичних моделях. Fear найчастіше реалізується через іменникове й клаузальне доповнення; anger характеризується прийменниковою спрямованістю та дієсловами вираження або регуляції; sadness профілюється через інтенсифікатори й зовнішньо сприймані ознаки; happiness входить до аксіологічних контекстів добробуту, прагнення й життєвої мети; disgust поєднує тілесну відразу з моральною оцінкою; surprise функціонує і як іменникова назва несподіваної події, і як показник її очікуваності або неочікуваності. Результати засвідчують, що COCA дає змогу описувати іменникові номінації емоцій не лише як словникові одиниці, а як контекстуально профільовані елементи мовної системи.*

***Ключові слова:*** *COCA, базові емоції, іменникові номінації емоцій, корпусна семантика, колокація, валентність*

## Introduction

A corpus-based study of emotional vocabulary must distinguish between emotion as an affective experience and the linguistic means by which that experience is named, evaluated, or construed in discourse. A corpus does not register emotion as an inner psychological process; it records recurrent choices of words and constructions through which emotional meaning becomes textually observable. The present analysis is therefore limited to six nominal labels of basic emotions - anger, fear, sadness, happiness, disgust, and surprise - and to their corpus behaviour in COCA.

The relevance of the study lies in the fact that emotional vocabulary is a part of the lexical system where nomination, evaluation, syntagmatic compatibility, and discourse function intersect. In contemporary English, nouns denoting emotions may indicate not only a state but also its source, intensity, addressee, outward manifestation, or evaluative framing. Corpus analysis enables observation of these features across recurrent contexts rather than in isolated illustrative examples.

The aim of the article is to substantiate the representativeness of COCA as a corpus source for the philological analysis of nominal labels of basic emotions in contemporary American English. To this end, the study examines the frequency,

collocational behavior, clusters, concordance environment, and grammatical-semantic patterns of the six lexemes. The analysis is concerned not with the psychological nature of emotions, but with the linguistic realization of the corresponding nominal labels in corpus material.

The unresolved issue for the present study concerns the integrated corpus description of the main nominal labels of basic emotions. Basic emotions have been widely discussed in psychology and affective science, while English emotion vocabulary has been studied through psycholinguistic norms, affective lexicons, bilingual lexical processing, historical semantics, discourse analysis, and computational approaches to emotion detection in text (Lindquist, 2021; Mohammad & Turney, 2013; Nandwani & Verma, 2021; Scott et al., 2019; Tissari, 2017; Warriner et al., 2013). What remains insufficiently described is the comparative profile of anger, fear, sadness, happiness, disgust, and surprise within a single corpus dataset, when frequency is interpreted alongside valency, collocation, cluster patterns, and concordance environments.

## Theoretical Background

The selection of the six lexemes follows Ekman's (1992) list of basic emotions. In this article, that model serves only as a principled criterion for selecting comparable nominal emotion labels; it is not used as evidence for the biological status of the corresponding emotions. This limitation is important because the theoretical status of basic emotions is disputed. Ortony and Turner (1990) argued that the category basic requires explicit classificatory criteria; Barrett (2006) questioned the treatment of emotions as natural kinds with stable boundaries; and Scherer (2005) proposed a componential model in which appraisal, bodily response, expression, action tendency, and subjective feeling are analytically distinct. For a corpus-based study, these positions establish a necessary methodological boundary: corpus data do not prove what an emotion is ontologically; they show how emotion meanings are represented in texts.

A second theoretical premise is the distinction between emotion-label words and emotion-laden words. Pavlenko (2008) distinguishes lexemes that directly name emotions from words that acquire affective force through experience, evaluation, cultural association, or context. Betancourt et al. (2024) further specify this distinction within a multicomponential account of affective meaning. In the present article, anger, fear, sadness, happiness, disgust, and surprise are treated as nominal emotion labels. Their collocates and concordance contexts, however, include emotion-laden units such as death, failure, violence, loss, pain, moral, profound,

paralyze, overcome, express, and control. The object of analysis is therefore not the noun in isolation, but the lexico-grammatical environment in which its meaning is profiled.

In contemporary philological research, emotional vocabulary is studied in journalistic, cinematic, media, phraseological, and digital discourse (Kots, 2023; Krysanova, 2019, 2023; Krysanova & Shevchenko, 2019; Krysanova & Verbytska, 2018; Mykhalchuk & Bihunova, 2019; Puhach, 2024; Verbytska & Krysanova, 2020; Zabuzhanska & Gregus, 2023; Zaluzhna & Maslona, 2023). These studies show that the linguistic representation of emotions depends not only on lexical meaning, but also on discourse environment, evaluation, modality, somatic signs, and culturally shaped models of interpretation.

The use of COCA in this study is motivated by its coverage of contemporary American English and by its register-balanced design. Following Biber (1993), representativeness in corpus design is not reducible to corpus size; it depends on the relation between the sampling principles and the linguistic object under investigation. For the present research, the decisive factor is that COCA's genre and chronological structure allow the lexico-semantic and contextual behaviour of nominal emotion labels to be compared across several types of contemporary American discourse.

Davies (2010) describes COCA as a monitor corpus created for the study of contemporary English usage across registers. In its current form, COCA contains approximately one billion words from texts covering 1990-2019 and includes spoken language, fiction, magazines, newspapers, academic texts, blogs, web pages, and film and television subtitles (Davies, n.d.). This design is relevant for the present topic because emotion labels do not function identically in conversational, fictional, journalistic, academic, and digital contexts.

Anokhina's (2019) dissertation is not focused on emotional vocabulary, yet it is methodologically relevant because it demonstrates the use of COCA, BNC, WebCorp, ParaSol, OPUS, Universal Dependencies, AntConc, SketchEngine, and other resources for corpus verification, clustering, valency analysis, distributional analysis, and concordance analysis. Fabian (2024) represents a related corpus-semantic orientation in the study of the English notion of power. In the present article, these works are significant not because their topics coincide with emotion, but because they illustrate the description of a lexical unit through frequency, co-occurrence, valency relations, and concordance evidence.

Recent COCA-based studies confirm the usefulness of this methodological perspective. Maniez (2017) shows that the spoken components of large corpora

require genre-sensitive interpretation. Sujatna et al. (2019) analyze modal auxiliary verbs through frequency and structural patterns. Tabbasum (2025) and Deliana and Gultom (2025) demonstrate how COCA can be used for synonym differentiation and collocational analysis. Although these studies do not examine nominal emotion labels directly, they show that semantic distinctions between lexemes can be described more reliably when frequency data are interpreted together with concordance evidence, collocational behaviour, and grammatical distribution.

**Methodological notes**

The study combines corpus-linguistic procedures, lexical-semantic interpretation, and the linguistic analysis of emotion. The corpus component supplies empirical evidence of attested usage: frequency, collocation, clusters, and concordance lines are treated as indicators of recurrent patterns in the functioning of the selected lexemes (McEnery & Hardie, 2012; Hunston, 2022). COCA was chosen because it is a large monitor corpus of contemporary American English and because its design corresponds to the object of this research: the lexico-semantic and contextual behaviour of six nominal emotion labels (Biber, 1993; Davies, 2010).

The lexical-semantic component concerns the interpretation of lexical meaning in use. The study assumes that the meaning of a noun is not exhausted by a dictionary definition; it is clarified by co-occurrence, valency, collocational environment, grammatical relations, and recurrent contexts. This position aligns with approaches that treat lexical meaning as a relation among the lexical unit, its paradigmatic and syntagmatic links, and its contexts of use (Cruse, 2011; Geeraerts, 2010). Accordingly, collocations are interpreted not as a mechanical list of neighbouring words, but as meaningful relations that reveal the semantic behaviour of the analysed emotion nouns (Brezina et al., 2015).

The linguistic analysis of emotion defines the selection and delimitation of the material. The nouns anger, fear, sadness, happiness, disgust, and surprise are treated as nominal emotion labels, that is, lexical units that directly name an emotional state. Their concordance environments often contain emotion-laden words that do not name an emotion directly but specify, intensify, or evaluatively modify it. The distinction between emotion-label words and emotion-laden words is central to the article because it separates direct emotion nomination from the broader affective vocabulary surrounding it in discourse (Pavlenko, 2008; Betancourt et al., 2024). This approach is also consistent with the philological study of English emotion words through semantic history, lexical relations, and contextual configurations (Tissari, 2017).

The study's material consists of COCA search results for six nouns: anger, fear, sadness, happiness, disgust, and surprise. The restriction to nominal emotion labels is deliberate: it prevents the analysis from conflating the selected nouns with adjectival predicates (e.g., *angry,* sad, happy, disgusted, and surprised), verbal forms (e.g., to fear and to surprise), or adverbial derivatives (e.g., angrily and surprisingly). The lexeme happiness is used as the nominal representative of the enjoyment/happiness domain, but its meaning is not equated with the full range of related nominations, including joy and pleasure. The analysis, therefore, remains focused on a single comparable grammatical class: nominal labels for basic emotions.

The corpus material was retrieved using the COCA tools Word, Topics, Collocates, Clusters, Texts/Virtual Corpora, and Concordance Lines. Word and Topics were used to determine frequency and topical associations. Collocates and Clusters were used to identify recurrent co-occurrences, phraseological tendencies, and repeated lexical sequences. Texts / Virtual Corpora and Concordance Lines were used to examine the selected nouns in attested contexts. This combination of tools enabled the interpretation of frequency alongside collocational behaviour, cluster patterns, and concordance evidence, which is necessary for a philological reading of corpus data.

The analysis was conducted in four stages. First, the absolute frequency of each noun was recorded and compared within the six-item sample. Second, topic associations, related words, collocates, and clusters were examined in order to define the nearest lexical environment of each unit. Third, recurrent patterns were grouped according to grammatical and semantic function: nominal and clausal complementation, prepositional government, intensification, coordination with other affective nouns, somatic or paralinguistic marking, and fixed expressions for evaluating expectedness. Fourth, these patterns were interpreted in terms of lexical valency, collocational behavior, semantic prosody, and the distinction between direct emotion nomination and affectively loaded contextual vocabulary.

The scope of the study is limited by the structure of the corpus and by the deliberately narrow lexical inventory. COCA is part-of-speech tagged and provides advanced search options, but it is not semantically annotated for emotion categories. Therefore, the lexical inventory was defined before the corpus search, and metaphorical, indirect, and idiomatic expressions of emotion were excluded from the main sample. Frequency counts are used as an initial empirical indicator, not as self-sufficient evidence of meaning. Since COCA represents American English, the findings cannot be automatically generalized to British, Australian, Indian, or other national varieties of English.

## Results and Discussion

The COCA data show that the six basic emotion labels are unevenly represented in the corpus. The total number of occurrences of *fear, surprise, anger, happiness, sadness*, and *disgust* in the analyzed material is 236,594. The frequency distribution is presented in Table 1. These figures do not reflect the actual intensity or prevalence of the corresponding emotions as mental states. They show how actively the selected nominal labels occur in the corpus texts. Accordingly, frequency is interpreted as an indicator of the lexical representation of emotional meanings in corpus material, not as direct evidence of the psychological nature of emotions or of how often they are experienced.

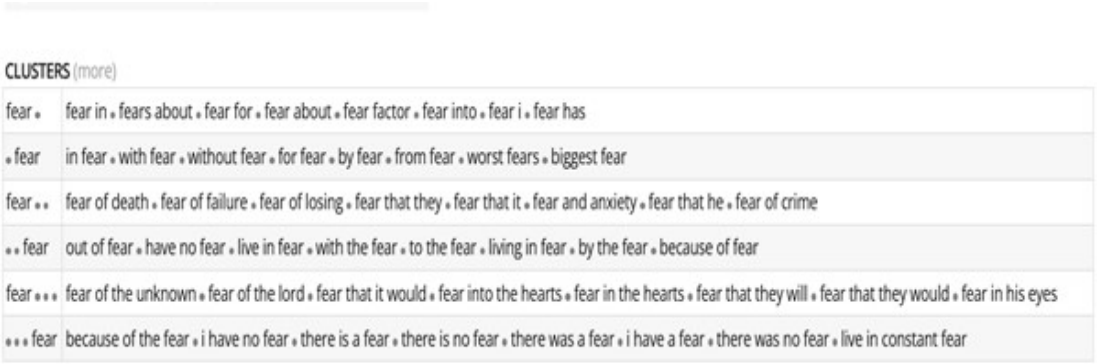
*Table 1.* Frequency distribution of six nominal emotion labels in COCA

| Lexeme | Frequency in COCA | Share within the six-item sample |
|---|---|---|
| Fear | 103,493 | 43.74% |
| Surprise | 60,286 | 25.48% |
| Anger | 36,471 | 15.42% |
| happiness | 22,057 | 9.32% |
| Sadness | 8,851 | 3.74% |
| Disgust | 5,436 | 2.30% |

The noun *fear* has the highest frequency in the analyzed sample. This predominance is associated with both the lexeme's direct nominative function and its broad semantic range. In COCA, *fear* is used in contexts of individual apprehension, social insecurity, religious discourse, political threat, risk, crime, possible failure, and potential harm. Its high frequency can therefore be related to its capacity to combine the naming of an emotion with the evaluation of danger, the anticipation of an undesirable outcome, and the expression of anxious expectation.

The noun *surprise* ranks second, but its frequency reflects a different type of use. The lexeme names an emotional reaction to an unexpected event and, at the same time, evaluates the relation between an event and the prior expectations of the speaker or addressee. This explains its frequent occurrence in fixed expressions such as *to my surprise* and *come as no surprise*, where the central feature is not emotional intensity but the relation between the event and the expected course of the situation. Disgust has the lowest frequency in the sample; nevertheless, its corpus combinations form a stable semantic profile. The lexeme tends to occur where aversion is represented through facial expression, gaze, voice, bodily movement, or moral-evaluative disapproval.

*Fear* regularly occurs in two grammatical-semantic patterns: *fear of + noun* and *fear that + clause*. In the first pattern, fear is associated with a named source of danger, such as an object, event, or condition: *fear of death, fear of failure, fear of losing, fear of crime, fear of the unknown*. In the second pattern, the content of apprehension is construed as a possible situation: *fear that they, fear that it, fear that he, fear that it would, fear that they will, fear that they would*. Thus, fear may denote both fear directed at a named object and apprehension about a possible development (Figure 1).



CLUSTERS (more)

| | |
|---|---|
| fear • | fear in • fears about • fear for • fear about • fear factor • fear into • fear i • fear has |
| • fear | in fear • with fear • without fear • for fear • by fear • from fear • worst fears • biggest fear |
| fear • • | fear of death • fear of failure • fear of losing • fear that they • fear that it • fear and anxiety • fear that he • fear of crime |
| • • fear | out of fear • have no fear • live in fear • with the fear • to the fear • living in fear • by the fear • because of fear |
| fear • • • | fear of the unknown • fear of the lord • fear that it would • fear into the hearts • fear in the hearts • fear that they will • fear that they would • fear in his eyes |
| • • • fear | because of the fear • i have no fear • there is a fear • there is no fear • there was a fear • i have a fear • there was no fear • live in constant fear |

*Figure 1*. Cluster patterns of the noun *fear* in COCA

The verbal collocates of fear clarify this interpretation. Overcome and allay present fear as a state that can be reduced or controlled. Strike and paralyze construe it as a force affecting the experiencer. Confirm connects fear with expectation, since a feared possibility may later be validated by an event. In the clusters *live in fear*, *living in fear, live in constant fear*, and *have no fear*, the noun denotes either a continuing state or the explicit absence of that state (Figure 2).

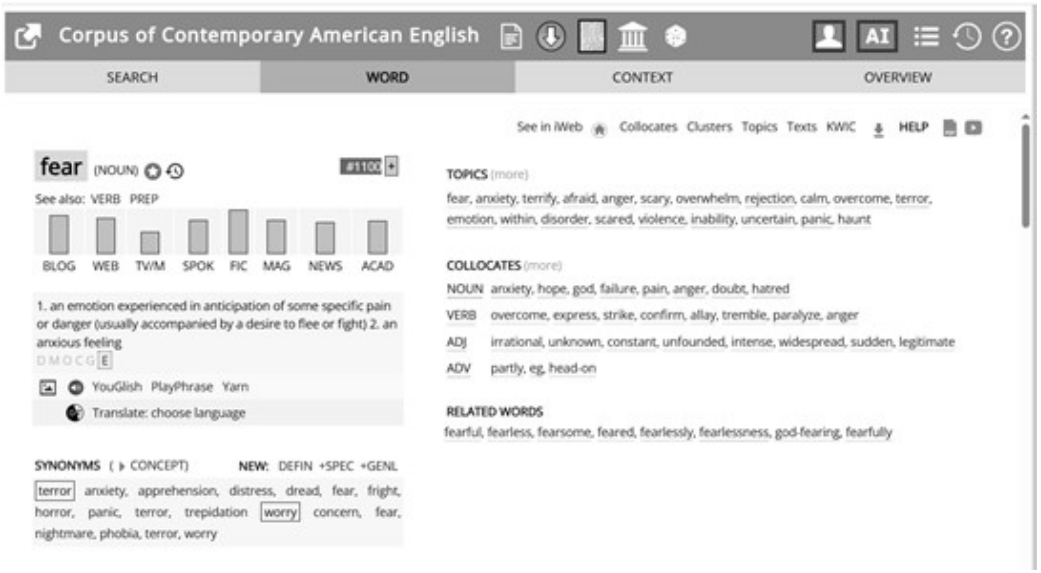
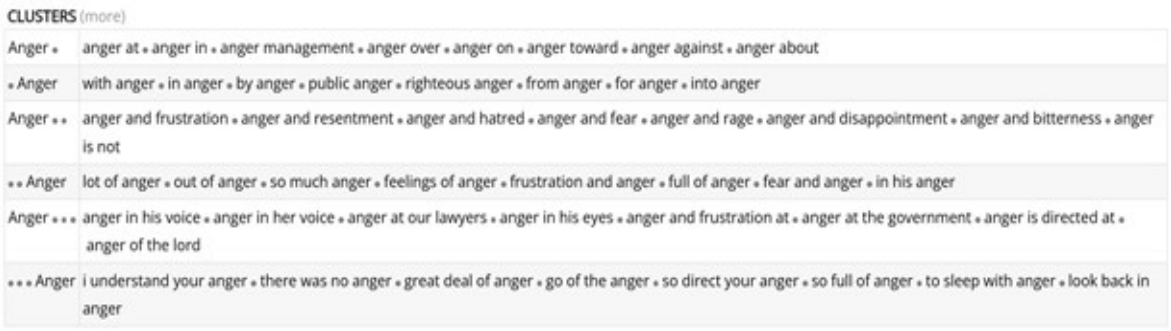



*Figure 2*. Collocational-semantic profile of the noun *fear* in COCA

The noun *anger* has a more explicitly directional valency profile than sadness or happiness. In the corpus material, it regularly occurs in the prepositional constructions *anger at, anger over, anger toward, anger against, and anger about*, which identify the cause, object, addressee, or adversarial orientation of negative evaluation (Figure 3).



| CLUSTERS (more) | |
|---|---|
| Anger • | anger at • anger in • anger management • anger over • anger on • anger toward • anger against • anger about |
| • Anger | with anger • in anger • by anger • public anger • righteous anger • from anger • for anger • into anger |
| Anger • • | anger and frustration • anger and resentment • anger and hatred • anger and fear • anger and rage • anger and disappointment • anger and bitterness • anger is not |
| • • Anger | lot of anger • out of anger • so much anger • feelings of anger • frustration and anger • full of anger • fear and anger • in his anger |
| Anger • • • | anger in his voice • anger in her voice • anger at our lawyers • anger in his eyes • anger and frustration at • anger at the government • anger is directed at • anger of the lord |
| • • • Anger | i understand your anger • there was no anger • great deal of anger • go of the anger • so direct your anger • so full of anger • to sleep with anger • look back in anger |

*Figure 3.* Cluster patterns of the noun *anger* in COCA

In combination with *at*, the noun *anger* most often refers to a specific person, action, event, or result that elicits a negative reaction. The construction *anger over* usually identifies an event, decision, or socially significant situation as the cause of anger. *Anger toward* marks the addressee of emotional directionality - a person, group, or social actor. By contrast, *anger against* strengthens the meaning of opposition and conflict because it indicates not only the cause of the emotion, but also its adversarial orientation. Thus, in corpus contexts, *anger* more often than *sadness* or *happiness* functions as a label for an emotion associated with an explicit external object of evaluation.

The verbal collocates of anger further specify its corpus profile. The verbs express, vent, direct, control, channel, fuel, and rise mark different facets of the emotional situation. Express and vent refer to verbal or outward manifestations; control and channel point to regulation and redirection; direct marks the addressee or target of the reaction; fuel denotes intensification caused by an additional factor; and rise denotes anger increasing over time. The expression anger management shows that anger may be conceptualized in the corpus as a state requiring conscious control, social regulation, or psychological correction. Adjectival combinations such as righteous anger, growing anger, sudden anger, intense anger, pent-up anger, and widespread anger specify ethical justification, gradual increase, sudden onset, intensity, accumulation, and collective distribution (Figure 4).


*Figure 4.* Collocational-semantic profile of the noun *anger* in COCA

The noun disgust also occurs in the prepositional constructions disgust at, disgust with, disgust for, and disgust over, but its semantic profile differs from that of anger. Whereas anger is usually directed toward a cause, addressee, or object of negative evaluation, disgust often combines evaluative rejection with somatic and paralinguistic marking. The clusters *look of disgust, disgust on his face, disgust in her voice, disgust in his eyes, turned away in disgust,* and *turn away in disgust* show that the noun is regularly associated with the face, gaze, voice, and bodily movement. In these contexts, the cluster environment specifies not only the cause of evaluation but also the visible or audible form in which aversion is expressed. The combination *moral disgust* extends the semantics of the noun from bodily aversion to moral-evaluative condemnation (Figure 5).

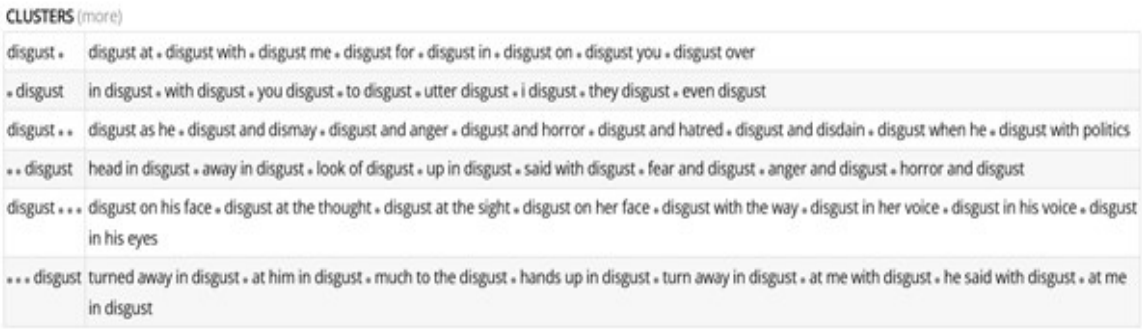
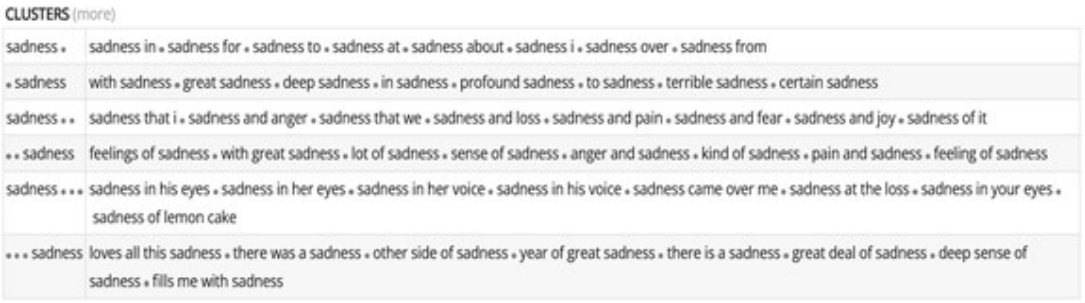



*Figure 5.* Cluster patterns of the noun *disgus*t in COCA

The noun *sadness* is less frequent than fear, surprise, anger, and happiness, but its corpus behaviour is internally consistent. The clusters *deep sadness, profound sadness, terrible sadness, overwhelming sadness, feelings of sadness, sense of*

*sadness, sadness in his eyes, sadness in her voice,* and *sadness came over me* point to two recurrent contextual patterns. The first is intensification through adjectives denoting depth, force, or scale. The second is the representation of sadness through perceptible signs such as gaze, voice, facial expression, or the general state of the speaker or character (Figure 6).



CLUSTERS (more)

| | |
|---|---|
| sadness • | sadness in • sadness for • sadness to • sadness at • sadness about • sadness i • sadness over • sadness from |
| • sadness | with sadness • great sadness • deep sadness • in sadness • profound sadness • to sadness • terrible sadness • certain sadness |
| sadness • • | sadness that i • sadness and anger • sadness that we • sadness and loss • sadness and pain • sadness and fear • sadness and joy • sadness of it |
| • • sadness | feelings of sadness • with great sadness • lot of sadness • sense of sadness • anger and sadness • kind of sadness • pain and sadness • feeling of sadness |
| sadness • • • | sadness in his eyes • sadness in her eyes • sadness in her voice • sadness in his voice • sadness came over me • sadness at the loss • sadness in your eyes • sadness of lemon cake |
| • • • sadness | loves all this sadness • there was a sadness • other side of sadness • year of great sadness • there is a sadness • great deal of sadness • deep sense of sadness • fills me with sadness |

*Figure 6*. Cluster patterns of the noun *sadness* in COCA

The contrast between *sadness* and *anger* is particularly informative. *Anger* is frequently realized through prepositional constructions that mark the direction of evaluation: *anger at, anger over, anger toward*. By contrast, sadness in the corpus material tends to combine with intensifiers or with the lexemes *eyes* and *voice*, which indicate outward signs of an internal state. This does not mean that *sadness* lacks causal motivation. Rather, the available contexts present sadness more often as the state of an experiencer than as a directed evaluative reaction to an external object.

Coordinative combinations such as *sadness* and *joy* require separate attention. They show that sadness should not be reduced to a purely negative emotional state. In narrative and reflective contexts, sadness may coexist with *joy, affection, memory, gratitude,* or *loss*, producing mixed affective configurations. The corpus material, therefore, supports an interpretation of sadness not only as a label for a negative state but also as a noun used in broader contexts of memory, loss, attachment, and evaluative reflection.

The noun disgust is also associated with externally perceptible signs, but their semantic function differs from that observed with sadness. In the combination *sadness in his eyes*, gaze helps identify an internal state. In contexts such as *disgust on his face or turned away in disgust,* facial expressions and bodily movements not only reveal the emotion but also convey a stance of rejection. Here, the bodily sign has an evaluative function: it represents not merely the presence of an emotional state, but a reaction of aversion. This explains why, in the corpus, disgust regularly combines with *horror, hatred, disdain, contempt,* and *moral judgment*, thereby

strengthening its semantics of aversion, condemnation, and socially or ethically motivated rejection.

The noun *happiness* differs from the other analyzed labels because much of its corpus environment belongs to the domains of value, well-being, and life purpose. The collocates *pursuit, health, love, joy, peace, success, project, satisfaction, and the verbs bring, pursue, wish, achieve, deserve, depend, sacrifice*, and *secure* indicate that *happiness* is often construed not as a momentary affective reaction, but as a desired state, a goal, a right, the result of personal or social effort, or a criterion of well-being (Figure 7).


*Figure 7.* Collocational-semantic profile of the noun *happine*ss in COCA

In the corpus material, *happiness* is therefore not limited to the naming of an immediate affective state. The lexeme regularly appears in contexts of life goals, moral judgment, interpersonal wishing, human rights, and the evaluation of quality of life. The combination of the *pursuit of happ*iness and *right to pursue happiness* refers to a recognizable formula of American political discourse, in which happiness functions as both a value and an object of aspiration. The expressions' *can't buy happiness'* and *'don't buy happiness'* formulate a moral-evaluative judgment about the limits of material prosperity. The expression *I wish you happiness* is an interpersonal wish in which *happiness* denotes a desired state of the addressee. Overall, happiness in COCA is connected with value, well-being, right, and life purpose, as well as with emotion.

The noun *surprise* is frequent for a different reason. The collocates *element, attack, visit, birthday, announcement, appearance, shock* and the clusters *surprise party* and *surprise attack* show that surprise may denote not only an emotional reaction, but also an event or situation defined by unexpectedness. The adjectival combinations *pleasant surprise*, *unpleasant surprise*, and *nasty surp*rise indicate that

the lexeme has no fixed positive or negative value; its evaluation depends on the nature of the event with which it is associated (Figure 8).

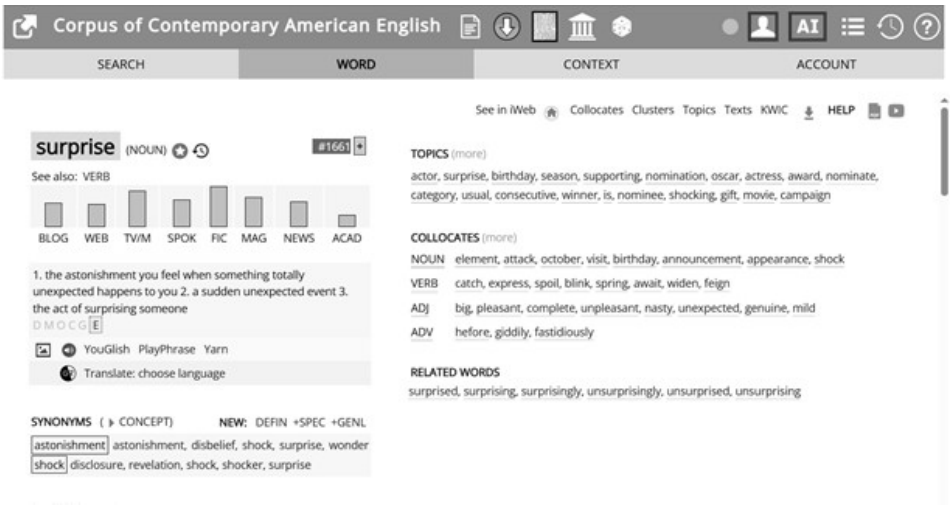



*Figure 8*. Collocational-semantic profile of the noun *surprise* in COCA

The fixed expressions with *surprise* are especially important: *to my surprise, to his surprise, come as a surprise, come as no surprise, it is no surprise, came as a surprise, it was a surprise, and it was no surprise*. In these contexts, the noun not only names an emotional reaction but also evaluates an event in terms of expectedness or unexpectedness. In *to my surp*rise, the speaker marks a mismatch between the event and prior expectations. By contrast, *'come as no surprise' or 'it is no surprise'* indicates that the event agrees with existing knowledge, assumptions, or predictions. In COCA, *surprise* therefore functions both as a nominal label for an emotional reaction and as a means of evaluating the predictability of an event.

Coordinative combinations help specify the semantic environment of each nominal label. *Anger* regularly combines with *frustration, resentment, fear, rage, disappointment, hatred,* and *bitterness,* which foreground dissatisfaction, conflict, emotional tension, and negative evaluation. *Fear* combines with *anxiety, doubt, pain, anger,* and *hatred,* linking them to uncertainty, threat, anticipation of undesirable developments, and psycho-emotional tension. *Sadness* combines with *loss, pain, fear, joy,* and *happiness;* consequently, in the corpus material, this label functions not only as a marker of a negative state but also as part of mixed affective configurations.

*Disgust* is coordinated with *anger, horror, hatred, disdain,* and *contempt,* which places it in the semantic area of aversion, condemnation, and evaluative rejection. *Happiness* combines with *joy, well-being, peace, satisfaction,* and *success*, forming a lexical environment associated with positive evaluation, harmony, and the attainment of a desired state. *Surprise* combines with *shock, announcement, appearance, attack,*

and *birthday*, which point to events, messages, sudden appearances, and unexpected outcomes.

In the present study, coordinative combinations are treated as evidence of the affective and evaluative environment of the analyzed nouns. *Anger* and *disgust* both belong to negative evaluation; however, *anger* is more often associated with dissatisfaction, conflictual reaction, and orientation toward an external object, whereas *disgust* emphasizes aversion, rejection, and morally or socially motivated condemnation. *Fear* and *sadness* may both denote an internal state of the experiencer, yet *fear* is oriented toward threat, uncertainty, and the anticipation of an undesirable outcome, while *sadness* is oriented toward loss, intensity, and perceptible signs of feeling. *Happiness* and *surprise* both occur in fixed expressions, but their functional orientations differ: *happiness* is connected with value, well-being, and life purpose, whereas *surprise* is connected with the evaluation of whether an event is expected or unexpected.

The comparison shows that the differences between the analyzed lexemes are not restricted to frequency or to positive and negative evaluation. Each nominal label has a distinct lexico-grammatical profile: some patterns involve prepositional government, others intensification, coordination with other affective nouns, fixed expressions, or contexts of external emotional manifestation. Table 2 summarizes the main grammatical-semantic patterns identified in the COCA material.

*Table 2.* Grammatical-semantic differentiation of six nominal emotion labels in COCA

| **Noun** | **Main pattern of co-occurrence** | **Representative COCA models** | **Philological interpretation** |
|---|---|---|---|
| fear | nominal and clausal complementation | fear of death; fear of failure; fear that...; live in fear | fear related to a source of danger or to a possible event |
| anger | prepositional directionality and verbs of expression/regulation | anger at; anger over; anger toward; express anger; control anger | negative evaluation directed toward a cause, addressee, or social object |
| sadness | intensification and externally perceptible signs | deep sadness; profound sadness; sadness in his eyes; sadness in her voice | an internal state intensified or inferred from visible or audible signs |
| happiness | contexts of value, well-being, and achievement | pursuit of happiness; true happiness; happiness and | a desired state, life goal, value, or |

| Noun | Main pattern of co-occurrence | Representative COCA models | Philological interpretation |
|---|---|---|---|
| | | well-being; can't buy happiness | criterion of well-being |
| disgust | somatic marking and evaluative rejection | look of disgust; disgust on his face; moral disgust; turned away in disgust | bodily aversion that may develop into moral-evaluative condemnation |
| surprise | naming an unexpected event and fixed expressions of expectedness | surprise party; surprise attack; to my surprise; come as no surprise | unexpectedness and the linguistic evaluation of an event's predictability |

The corpus data point to several regularities. First, the analyzed nominal labels differ in grammatical co-occurrence: *fear* combines with nominal and clausal complementation; *anger* and *disgust* tend toward prepositional constructions; sadness is often realized through intensifiers and perceptible signs; *happiness* occurs in contexts of value, well-being, and achievement; and *surprise* is regularly used in fixed expressions that evaluate expectedness or unexpectedness. Evaluative semantics also differentiates the lexemes: *anger* and *disgust* both belong to negative evaluation, but anger is primarily directed toward the cause or object of reaction, whereas disgust is associated with aversion, rejection, and condemnation. Frequency alone does not account for the semantic profile of the nouns: the less frequent *disgust* has a sharply defined collocational environment, while the highly frequent *surprise* includes both an emotional reaction proper and linguistic evaluation of event predictability.

For this reason, COCA is useful for philological analysis of emotional vocabulary: it allows the researcher to describe not emotion itself as an inner mental state, but the lexical, grammatical, and contextual forms through which emotion is represented in text. The object of analysis is therefore not the psychological nature of emotion, but the corpus behaviour of its nominal label.

## Conclusions and Prospects for Further Research

The analysis demonstrates that the nominal labels *anger, fear, sadness, happiness, disgust*, and *surprise* differ not only in frequency in COCA, but also in lexico-grammatical and semantic behaviour. Fear has the highest frequency and occurs in the patterns *fear of + noun* and *fear that + clause*, which relate the lexeme respectively to a source of danger and to a possible course of events. *Surprise* ranks second in frequency and functions as a label for an emotional reaction, a name for an

unexpected event, and a means of evaluating whether that event was expected or unexpected.

*Anger* is characterized by prepositional directionality toward a cause, addressee, or object of negative evaluation, while its verbal collocates associate this label with expression, regulation, orientation, and intensification. In the corpus material, *sadness* is more often realized through intensifiers and perceptible signs; *happiness* occurs in contexts of value, well-being, life purpose, and achievement; and *disgust* combines bodily aversion with moral-evaluative condemnation. Basic emotions are therefore treated in this article not as psychological entities, but as nominal labels whose corpus profiles are shaped by recurrent co-occurrence, grammatical relations, collocational environment, clusters, and concordance contexts.

COCA is appropriate for this type of analysis because its register-balanced corpus structure allows frequency counts to be compared with the contexts in which emotional vocabulary functions. At the same time, frequency in this study does not indicate the prevalence of emotional states in extralinguistic reality; it indicates the degree of lexical representation of the corresponding labels in contemporary American English.

Further research may proceed in three directions. First, the analysis can be extended from individual nominal labels to lexical families, including adjectival, verbal, and adverbial forms such as angry, angrily, afraid, fearful, scared, sad, sadly, happy, happily, disgusted, disgusting, surprised, and surprising. This would show how one emotional domain is represented across different parts of speech. Second, genre comparison within COCA is promising, since fiction, spoken language, academic texts, news, magazines, blogs, web pages, and subtitles may differ in their use of intensification, evaluation, and discourse functions of emotion labels. Third, comparison with the British National Corpus, GloWbE, or specialized corpora of digital discourse would help distinguish American patterns of emotional vocabulary from broader English-language tendencies.

## *REFERENCES*

1. Anokhina, T. O. (2019). *Typolohiia lakunikoniv anhlomovnoi ta ukrainomovnoi kartyn svitu* [Typology of lacunicons of the English- and Ukrainian-language worldviews] (Extended abstract of Doctoral dissertation). National Pedagogical Dragomanov University. https://old.npu.edu.ua/images/file/vidil_aspirant/avtoref/D_26.053.26/Anokhina.pdf (in Ukrainian)
2. Barrett, L. F. (2006). Are emotions natural kinds? *Perspectives on Psychological Science*, 1(1), 28–58. https://doi.org/10.1111/j.1745-6916.2006.00003.x
3. Betancourt, Á.-A., Guasch, M., & Ferré, P. (2024). What distinguishes emotion-label words from emotion-laden words? The characterization of affective meaning from a multi-componential

conception of emotions. *Frontiers in Psychology*, 15, Article 1308421. https://doi.org/10.3389/fpsyg.2024.1308421

4. Biber, D. (1993). Representativeness in corpus design. *Literary and Linguistic Computing*, 8(4), 243–257. https://doi.org/10.1093/llc/8.4.243

5. Brezina, V., McEnery, T., & Wattam, S. (2015). Collocations in context: A new perspective on collocation networks. *International Journal of Corpus Linguistics*, 20(2), 139–173. https://doi.org/10.1075/ijcl.20.2.01bre

6. Cruse, A. (2011). Meaning in language: An introduction to semantics and pragmatics (3rd ed.). *Oxford University Press*.

7. Davies, M. (2010). The Corpus of Contemporary American English as the first reliable monitor corpus of English. *Literary and Linguistic Computing*, 25(4), 447–464. https://doi.org/10.1093/llc/fqq018

8. Davies, M. (n.d.). The Corpus of Contemporary American English (COCA). *English-Corpora.org.* Retrieved May 10, 2026, from https://www.english-corpora.org/coca/

9. Deliana, & Gultom, A. F. (2025). Collocations of effective and efficient in COCA: A corpus analysis. *International Journal of Culture and Art Studies*, 9(2), 117–126. https://doi.org/10.32734/ijcas.v9i2.22811

10. Ekman, P. (1992). An argument for basic emotions. *Cognition & Emotion*, 6(3–4), 169–200. https://doi.org/10.1080/02699939208411068

11. Fabian, M. (2024). Corpus-based methods of power notion analysis in English. *Contemporary Studies in Foreign Philology*, 1(25), 196–208. https://doi.org/10.32782/2617-3921.2024.25.196-208

12. Geeraerts, D. (2010). Theories of lexical semantics. *Oxford University Press*. https://doi.org/10.1093/acprof:oso/9780198700302.001.0001

13. Hunston, S. (2022). Corpora in applied linguistics (2nd ed.). *Cambridge University Press*. https://doi.org/10.1017/9781108616218

14. Kots, T. E. (2023). Emotyvni mekhanizmy suchasnykh publitsystychnykh tekstiv [Emotive mechanisms of modern publicistic texts]. *Linhvistyka*, 2(48), 98–106. https://doi.org/10.12958/2227-2631-2023-2-48-98-106 (in Ukrainian)

15. Krysanova, T. (2019). Constructing negative emotions in cinematic discourse: A cognitive-pragmatic perspective. *Cognition, Communication, Discourse*, 19, 55–77. https://doi.org/10.26565/2218-2926-2019-19-04

16. Krysanova, T. (2023). Psycholinguistic and cognitive-semiotic dimensions of constructing fear in horror films: A multimodal perspective. *East European Journal of Psycholinguistics,* 10(1), 96–115. https://doi.org/10.29038/eejpl.2023.10.1.kry

17. Krysanova, T., & Shevchenko, I. (2019). The intersemiosis of negative emotions in the cinematic discourse: A psycholinguistic perspective. *Psycholinguistics*, 25(2), 117–137. https://doi.org/10.31470/2309-1797-2019-25-2-117-137

18. Krysanova, T., & Verbytska, A. (2018). Pryntsypy linhvistychnoho konstruyuvannya emotsiy u dyskursi [Principles of linguistic construction of emotions in discourse]. *Linhvistychni Studiyi*, 36, 78–84. https://doi.org/10.31558/1815-3070.2018.36.8 (in Ukrainian)

19. Lindquist, K. A. (2021). Language and emotion: Introduction to the special issue. *Affective Science*, 2, 91–98. https://doi.org/10.1007/s42761-021-00049-7

20. Maniez, F. (2017). Representation of conversational style in the oral components of the BNC and the COCA: Towards the description of a mixed genre. *Recherches anglaises et nord-américaines*, 50, 91–106. https://doi.org/10.3406/ranam.2017.1551

21. McEnery, T., & Hardie, A. (2012). Corpus linguistics: Method, theory and practice. *Cambridge University Press*. https://doi.org/10.1017/CBO9780511981395

22. Mohammad, S. M., & Turney, P. D. (2013). Crowdsourcing a word-emotion association lexicon. *Computational Intelligence*, 29(3), 436–465. https://doi.org/10.1111/j.1467-8640.2012.00460.x

23. Mykhalchuk, N., & Bihunova, S. (2019). The verbalization of the concept of fear in English and Ukrainian phraseological units. *Cognitive Studies / Études cognitives*, 19, Article 2043. https://doi.org/10.11649/cs.2043

24. Nandwani, P., & Verma, R. (2021). A review on sentiment analysis and emotion detection from text. *Social Network Analysis and Mining*, 11, Article 81. https://doi.org/10.1007/s13278-021-00776-6

25. Ortony, A., & Turner, T. J. (1990). What's basic about basic emotions? *Psychological Review*, 97(3), 315–331. https://doi.org/10.1037/0033-295X.97.3.315

26. Pavlenko, A. (2008). Emotion and emotion-laden words in the bilingual lexicon. *Bilingualism: Language and Cognition*, 11(2), 147–164. https://doi.org/10.1017/S1366728908003283

27. Puhach, Y. M. (2024). The verbalization of emotions in English phraseological units. *Scientific Journal of Dragomanov Ukrainian State University. Series 9. Current Trends in Language Development*, 27, 38–50. https://doi.org/10.31392/NPU-nc.series9.2024.27.04

28. Scherer, K. R. (2005). What are emotions? And how can they be measured? *Social Science Information*, 44(4), 695–729. https://doi.org/10.1177/0539018405058216

29. Scott, G. G., Keitel, A., Becirspahic, M., Yao, B., & Sereno, S. C. (2019). The Glasgow Norms: Ratings of 5,500 words on nine scales. *Behavior Research Methods*, 51, 1258–1270. https://doi.org/10.3758/s13428-018-1099-3

30. Sujatna, M. L., Sujatna, E. T. S., & Pamungkas, K. (2019). Exploring the use of modal auxiliary verbs in Corpus of Contemporary American English (COCA). *Sosiohumaniora,* 21(2), 166–172. https://doi.org/10.24198/sosiohumaniora.v21i2.19970

31. Tabbasum, R. (2025). A corpus-based comparative study of the verbs suggest, recommend and propose in the Corpus of COCA. *CORPORUM: Journal of Corpus Linguistics*, 8(1), 16–24.

32. Tissari, H. (2017). Current emotion research in English linguistics: Words for emotions in the history of English. *Emotion Review*, 9(1), 86–94. https://doi.org/10.1177/1754073916632064

33. Verbytska, A., & Krysanova, T. (2020). DISTRESS in English media: Integrating cognitive-discursive and computational approaches. *Cognition, Communication, Discourse,* 21, 72–92. https://doi.org/10.26565/2218-2926-2020-21-06

34. Warriner, A. B., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods*, 45, 1191–1207. https://doi.org/10.3758/s13428-012-0314-x

35. Zabuzhanska, I., & Greguš, Ľ. (2023). Emotional storytelling technique in English and Slovak media discourse: Points of linguistic intersection. *Lege Artis. Language Yesterday, Today, Tomorrow*, 8(2), 94–109. https://doi.org/10.34135/lartis.23.8.2.06

36. Zaluzhna, O., & Maslona, V. (2023). Verbalization of emotional states in modern English spoken Internet discourse: On the material of video reviews of cosmetic products. *Scientific Notes. Series: Philological Sciences*, 3(206), 41–46. https://doi.org/10.32782/2522-4077-2023-206-6

The author declares no competing interests.



Publisher: Borys Grinchenko Kyiv Metropolitan University, Kyiv, Ukraine

Publishing history: Received 29.30.2025 ; Accepted 12.05.2026; Published 30.05.2026