

BNC AND COCA AS CORPUS RESOURCES OF THE VERBAL EMOTICON IN CONTEMPORARY ENGLISH: REPRESENTATIVENESS, SEARCH TOOLS, AND LIMITS OF INTERPRETATION

Nataliia Bober

PhD, Associate Professor, Associate Professor at the Department of Germanic Philology,
Faculty of Romance and Germanic Philology,
Borys Grinchenko Kyiv Metropolitan University, Ukraine
e-mail: n.bober@kubg.edu.ua, orcid.org/0000-0002-9639-0562

Rusudan Makhachashvili

Dr. Hab., Full Professor,
Head of Internationalization of Higher Education Research Lab,
Borys Grinchenko Kyiv Metropolitan University, Ukraine
e-mail: r.makhachashvili@kubg.edu.ua, orcid.org/0000-0002-4806-6434

Summary

The article substantiates the methodological value of the British National Corpus and the Corpus of Contemporary American English for the study of the verbal emoticon in contemporary English. The verbal emoticon is defined as a functional-discursive category: a verbal unit or construction that names, frames, intensifies, evaluates, or performs emotion in a specific context. The category includes lexical nominations, predicative patterns, evaluative exclamatives, interjective formulas, metaphorical expressions, and phraseological units when they function as markers of affective stance or interpersonal alignment. The study argues that BNC and COCA should not be used as interchangeable sources. BNC provides documented British evidence and enables comparison between written and spoken material, while COCA offers a large American corpus with genre-sensitive access to frequency, KWIC, collocation, and diachronic distribution. The article develops a corpus-based procedure that combines inventory building, concordance reading, collocational and constructional analysis, genre comparison, and contextual annotation. It also defines the interpretive limits of corpus evidence: frequency measures textual representation rather than psychological occurrence; collocation indicates recurrent co-selection but does not explain meaning without context; and corpus architecture determines the scope of possible conclusions. The final section outlines the relevance of corpus-grounded semantic profiling for large language models, especially for emotion recognition, stance detection, and affect-sensitive language generation.

Key words: affective stance; concordance; collocation; discourse pragmatics; emotional language; semantic profiling; language models.

DOI <https://doi.org/10.23856/7606>

1. Introduction

The corpus study of emotion-marked language is productive only when the researcher treats corpus evidence as interpretable linguistic data rather than as a substitute for interpretation. Emotional meaning in contemporary English is not confined to dictionary definitions of emotion words. It is distributed across repeated lexical choices, grammatical patterns, formulaic

reactions, genre conventions, and pragmatic positions. A corpus can register these recurrences, but it cannot decide by itself whether a given occurrence names an emotion, performs a reaction, frames an evaluation, or signals interpersonal alignment. This distinction is central to the present article.

The **study object** is the verbal emoticon. In this article, the term does not denote a graphic emoticon, emoji, or typographic sign. It refers to a set of verbal units and constructions through which a speaker transforms emotion into an experienced state, an affective reaction, an evaluative stance, or an interpersonal orientation. The category may include nouns, adjectives, adverbs, verbs, interjections, phraseological units, metaphorical predicates, and syntactic constructions, provided that in a concrete context they perform affective or stance-taking work. The noun *relief* and the construction *What a relief!* are therefore not equivalent objects: the first may name an emotional state, while the second performs an affective response. The same logic applies to: *"I am absolutely thrilled."* *Seriously?* *That is disgusting!*, and *That made my day*.

The **relevance of the topic** follows from a methodological gap in corpus-based studies of emotional language. English emotion vocabulary has been analysed in semantics, pragmatics, cognitive linguistics, corpus linguistics, and computational linguistics, yet the verbal emoticon as a broader functional-discursive category has not been sufficiently described. *Bednarek (2008)* shows that emotion talk varies across corpora and genres; *Pavlenko (2008)* distinguishes words that name emotions from words that become affective through experience and discourse; *Wierzbicka (1999)* demonstrates that emotion concepts are linguistically and culturally structured; *Mohammad and Turney (2013)* operationalize word-emotion associations for computational use. These approaches are necessary, but the verbal emoticon requires an additional step: the researcher must explain how different formal resources converge on a single discourse function.

The article focuses on two corpus sources: the British National Corpus (BNC) and the Corpus of Contemporary American English (COCA). They are frequently mentioned together in English linguistic research, but they do not provide identical evidence. BNC offers a documented model of British English, with written and spoken components and stable metadata. COCA provides a large American corpus with access to genre distribution, frequency, KWIC lines, collocates, and cross-time comparisons. The question is not which corpus is superior. The relevant question is what each corpus allows the researcher to claim about the verbalization of emotion.

The **aim of the article** is to justify BNC and COCA as corpus sources for the study of the verbal emoticon in contemporary English and to propose a procedure for selecting, verifying, and interpreting corpus material. The objectives are: to define the verbal emoticon as a corpus-relevant category; to compare BNC and COCA in terms of representativeness, architecture, and search potential; to describe the analytical procedures needed for identifying verbal emoticons; to explain how semantic profiling can be based on concordance and collocational evidence; to determine the limits of corpus interpretation; and to indicate how the results may inform research on large language models.

The **novelty of the study** lies in treating the choice of corpus as part of the theory of the object, not as a merely technical step. BNC and COCA are interpreted as differently structured sources through which the same emotion-marked form may receive different frequencies, genre positions, collocational profiles, and pragmatic functions. This view follows corpus design theory, in which representativeness depends on sampling, the target population, text categories, and research purpose rather than on corpus size alone (*Atkins, Clear, & Ostler, 1992; Biber, 1993; Biber, Conrad, & Reppen, 1998*). It also follows the methodological principle that

corpus linguistics is a research practice requiring explicit procedures, not a technology that delivers conclusions automatically (McEnery & Hardie, 2012).

2. Verbal Emoticon and Corpus Evidence

The verbal emoticon is best understood as a functional category. Its defining property is not part of speech, length, or graphic form, but the discourse work performed by a unit in context. This understanding prevents two reductions. The first reduction identifies verbal emoticons only with interjections and short reactions such as *Wow!*, *Oh no!*, or *Yay!*. The second treats every emotion word as a verbal emoticon regardless of use. Both approaches are insufficient. An interjection may be purely phatic in one context, while an emotion noun may occur in a diagnostic or institutional register without performing interpersonal affect.

The category becomes coherent when described in terms of function. A verbal emoticon participates in one or more of four operations: representing emotion, evaluating a situation, marking stance, or aligning the speaker with an interlocutor. These functions often overlap. *I am so happy for you* expresses a personal affective state, evaluates the addressee's situation positively, and creates interpersonal solidarity. *That is awful!* may express emotional reaction and moral judgement at the same time. *Seriously?* may encode surprise, disbelief, irritation, or scepticism, depending on the preceding utterance and the genre of interaction.

This functional interpretation is compatible with existing research but does not duplicate it. Hunston (2007) argues that semantic prosody emerges from recurrent contexts, not from isolated words. Sinclair (1991) shows that meaning is often organized through co-selection. Tognini-Bonelli (2001) distinguishes corpus-based analysis, in which the researcher starts with a prior category, from corpus-driven analysis, in which the category emerges from the data. The present study is corpus-based because it begins with the verbal emoticon as a philological category; however, the category remains open to correction in light of concordance evidence.

Ukrainian corpus and digital-linguistic scholarship are also relevant here. Demska-Kulchytska (2005) defines corpus work through the selection, organization, and annotation of linguistic evidence. Holoshchuk (2021) treats corpus linguistics as a practical instrument of contemporary philology, while Zhukovska (2023) demonstrates the value of quantitative corpus-based methods for construction grammar. Anokhina (2018) is methodologically important because her work on a corpus lacunicon insists on precise selection criteria for machine-readable lexical material. These positions support the present argument that corpus evidence must be organized before it can be interpreted.

The author's earlier dissertation on matrix profiling of phrasal-verb emotives in the British National Corpus provides a direct methodological precedent for this article. Bober (2020) developed a model of cognitive-semantic submatrices for phrasal-verb emotives in BNC and showed that emotional meanings can be reconstructed through corpus queries, concordance evidence, statistical profiling, and semantic clustering. The present article extends that logic from phrasal-verb emotives to a broader category that includes lexical, constructional, interjective, metaphorical, and evaluative realizations of affective stance.

The distinction between verbal emoticons and graphic emoji is equally important. Bakhtina (2023) conceptualizes emoji in computer being as a polylateral sign whose meaning depends on generation, semanticization, visualization, and pragmatic interpretation in digital communication. Makhachashvili (2022) analyses language development in the digital environment through the interaction of linguistic models and communicative technologies. These studies justify the boundary that the present article preserves: emoji and verbal emoticons may

serve related affective functions, but they are distinct semiotic resources. The present study concerns verbal units that can be retrieved and interpreted in corpora.

3. BNC and COCA as Resources

Representativeness is decisive for any corpus claim. A corpus does not represent a language in the abstract; it represents a designed sample of language use. Its evidential value depends on period, genre, register, national variety, annotation, metadata, and access interface. This principle is especially important for emotional language because affective expression is sensitive to communicative conditions. A formula frequent in dialogue may be marginal in academic prose; an evaluative pattern common in web writing may be absent from a late twentieth-century corpus; a metaphorical reaction may occur differently in British and American evidence.

The British National Corpus remains a central reference corpus for British English. The BNC XML Edition contains approximately 100 million words and is supported by documented metadata, lemmatization, part-of-speech tagging, and text classification (*BNC Consortium, 2007; British National Corpus Consortium, 2007*). For the study of verbal emoticons, its value lies not in maximal size but in the possibility of examining emotion-marked units against a structured sample of British written and spoken language. BNC1994 should not be treated as a direct representation of present-day digital English. Its value lies in its ability to reconstruct the pre-social-media background of verbal mechanisms, now intensified in digital discourse.

The BNC2014 project is relevant because it updates the British corpus tradition. The Spoken BNC2014 records everyday conversations from the 2010s (*Love, Dembry, Hardie, Brezina, & McEnery, 2017*). The Written BNC2014 was designed as a contemporary counterpart to the written component of BNC1994 and is useful for diachronic comparability (*Brezina, Hawtin, & McEnery, 2021*). The distinction matters because emotional routines in spontaneous conversation, public written discourse, and online-oriented writing cannot be interpreted as the same type of evidence.

COCA responds to a different research need. *Davies (2010)* presents COCA as a reliable monitor corpus of American English, and *Davies (2020)* describes its architecture as a large, genre-diverse resource with spoken language, fiction, magazines, newspapers, academic texts, TV and movie subtitles, blogs, and web pages. For verbal emoticon research, these components are not merely categories of storage. They are discourse environments in which affective stance is differently produced. Fiction may construct character psychology; subtitles may approximate dialogue; news may frame public affect; web and blog texts may preserve newer evaluative formulas.

The comparison of BNC and COCA must not be reduced to a simple British-American opposition. *Kilgarriff (2001)* warns that corpora cannot be compared by impression because their internal composition influences which linguistic features become statistically prominent. A higher frequency of a form in COCA may reflect American usage, genre distribution, period coverage, or interface design. A lower frequency in BNC may reflect historical composition rather than linguistic absence. Responsible comparison, therefore, requires attention to corpus architecture before interpretation.

4. Methodology of Identification

The material of the study is a corpus-searchable inventory of verbal units and constructions that may perform an affective or stance-taking function. The initial inventory includes direct emotion nominations such as fear, anger, sadness, joy, disgust, and surprise; adjectival and verbal markers such as afraid, angry, happy, disgusted, surprised, enjoy, hate, and grieve;

recurrent patterns such as fear of, fear that, anger at, disgust with, be afraid, feel happy, and be surprised at; interjective and evaluative formulas such as *Oh no!*, *Wow!*, *What a relief!*, *I cannot believe this!*, and *That is disgusting!*; and metaphorical expressions such as *That made my day* and *That broke my heart*. Ekman's (1992) model of basic emotions helps justify the initial set of direct nominations, but the corpus procedure is not limited to basic-emotion theory.

The first methodological operation is inventory building. The researcher defines searchable forms and patterns, records the query type, and distinguishes word, lemma, phrase, and constructional searches. The second operation is frequency profiling. Frequency is recorded together with corpus version, genre, query form, and date of extraction. Gries (2017) and Brezina (2018) emphasize that statistical values must be interpreted through sampling, normalization, and distribution; this is indispensable when corpora differ in size and design.

The third operation is concordance reading. KWIC lines and expanded contexts determine whether an occurrence performs the relevant function. Fear in *fear of crime* may be a public concern rather than a private emotion. Anger in *public anger* over policy may denote collective evaluation. *Disgust* may refer to sensory aversion, moral condemnation, or political judgement. *Surprise* may indicate emotion, unexpectedness, or discourse management. These distinctions cannot be inferred from frequency alone.

The fourth operation is collocational and constructional analysis. Collocates such as *deep*, *sudden*, *growing*, *public*, *moral*, *bitter*, *genuine*, and *overwhelming* specify intensity, temporality, scope, evaluation, and discourse orientation. Prepositional patterns also matter: *fear of*, *fear that*, *anger at*, *anger over*, *disgust with*, and *surprise at* distribute semantic roles differently. The collocational approach developed by Stefanowitsch and Gries (2003) and extended by Gries and Stefanowitsch (2004) is useful because it identifies how words and constructions co-select one another.

The fifth operation is contextual annotation. Each relevant occurrence can be classified as direct emotion naming, state attribution, affective reaction, evaluative stance, semantic prosody, or interpersonal alignment. These categories are diagnostic, not rigid. One occurrence may combine several functions. This is why the annotation must remain philological and cannot be replaced by automatic labelling.

Table 1

Corpus procedure for identifying verbal emoticons

Corpus operation	Corpus material	Diagnostic question	Analytical result
Inventory building	fear, thrilled, Seriously?, What a relief!	Which forms may perform affective or stance-taking work?	Searchable inventory
Concordance reading	KWIC lines and expanded contexts	Does the occurrence name, perform, evaluate, or align emotion?	Exclusion of irrelevant uses
Collocational analysis	deep fear, public anger, absolutely thrilled	What recurrent semantic environment surrounds the form?	Semantic preference and prosody
Genre comparison	spoken, fiction, web, news, and academic texts	Where does the function become discourse-relevant?	Register-sensitive interpretation
Contextual annotation	naming, state, reaction, stance, alignment	What discourse role is performed?	Functional classification

5. Semantic Profiling

Semantic profiling reconstructs the recurrent semantic, grammatical, and pragmatic behaviour of a form or construction. In the study of verbal emoticons, a profile includes typical collocates, grammatical patterns, genre distribution, pragmatic function, and affective orientation. The profile of fear may include object-oriented constructions such as *fear of*, propositional constructions such as *fear that*, modifiers such as *deep* or *growing*, and institutional contexts such as *public fear* or *fear of crime*. Such evidence describes the textual organization of emotion; it does not describe psychological experience.

Direct emotion nominations constitute only the first layer of the profile. Words such as *fear*, *anger*, *sadness*, *joy*, *disgust*, and *surprise* are searchable and theoretically stable, but they do not exhaust the object. *Anger* may be private, public, political, moral, or rhetorical. *Disgust* may connect bodily aversion with ethical judgment. *Surprise* may refer to affective astonishment or to unexpectedness without emotional salience. *Joy* may appear as an abstract noun or as part of an exclamative construction.

State predicates form a second layer. Expressions such as *I am afraid*, *she was angry*, *they felt happy*, and *he was disgusted* attribute emotion to a subject. Their profile depends on subject type, complement structure, intensification, and genre. A first-person predicate may perform self-disclosure; a third-person predicate may describe character or report another person's emotion; an institutional subject, as in the committee, is concerned, may produce a cautious administrative stance rather than an affective confession.

Reaction formulas form a third layer. *Oh no!*, *Wow!*, *I cannot believe this!*, *What a relief!*, and *That is awful!* often perform emotion instead of describing it. In digital discourse, they may compensate for the absence of intonation, facial expression, and immediate co-presence. In corpora they require careful treatment because punctuation, capitalization, and transcription conventions influence retrieval.

Metaphorical and idiomatic expressions form a fourth layer. *That made my day*, *That broke my heart*, *I was blown away*, and *It drives me mad* show that verbal emoticons may encode emotion indirectly. Their meanings are often conventional enough to be searchable, but context remains decisive. A phrase may be literal in one context and affective in another. Corpus evidence reveals recurrent use; interpretation determines the function.

The profile of a verbal emoticon can be organized through four axes: emotional representation, evaluation, stance, and interpersonal alignment. Emotional representation concerns naming or attributing feeling. Evaluation concerns approval, disapproval, relief, disgust, admiration, disappointment, or irritation. Stance concerns the speaker's position toward propositional content. Interpersonal alignment concerns solidarity, empathy, distance, irony, or support. This model explains why formally distinct expressions may belong to the same category when they perform the same affective-pragmatic work.

6. Limits of Interpretation

The first limit of corpus interpretation is the difference between textual representation and psychological occurrence. A corpus shows how often a word, phrase, or construction occurs in selected texts. It does not show how often speakers experience the corresponding emotion. A high number of fear tokens in a corpus indicates textual visibility within that corpus, not a higher amount of fear in the speech community. This distinction prevents a common methodological error: treating frequency as emotional reality.

The second limit is polyfunctionality. Emotion-marked forms move between subjective state, social evaluation, public discourse, irony, and metaphorical extension. Concern may indicate affect, institutional caution, or policy framing. Surprise may express feeling, unexpectedness, or conversational management. Disgust may be sensory, moral, political, or rhetorical. Annotation must therefore allow functional ambiguity instead of forcing every instance into a single label.

The third limit is multimodality. Verbal emoticons in digital discourse often occur with emojis, punctuation, capitalization, images, GIFs, reaction buttons, or layout features. Traditional corpora preserve these resources unevenly. *Bakhtina's (2023)* analysis of emoji as a polylateral sign helps explain why the verbal component may not exhaust the affective meaning of a digital message. A corpus may contain *Seriously?* but not the visual cue that changes it from disbelief into humour or irritation.

The fourth limit concerns corpus design and interface. BNC1994 is historically valuable but temporally bounded. BNC2014 updates parts of the British evidence, but should be specified by component. COCA is large and genre-rich but represents American English through its own categories. Search tools also influence results: word queries differ from lemma queries, phrase queries may miss intervening modifiers, and punctuation-sensitive forms may be inconsistently represented. A publication must therefore state corpus version, query form, exclusion criteria, genre scope, and annotation principles.

These limits do not weaken corpus research. They define the conditions under which corpus evidence can support a responsible philological conclusion. *Hunston (2002)* is useful here because applied corpus work depends not on extracting examples but on explaining what corpus tools can and cannot show. The verbal emoticon is a context-dependent object, so corpus analysis must remain both empirical and interpretive.

7. Implications for Large Language Models

The corpus study of verbal emoticons is directly relevant to large language models. Transformer architecture made it possible to model contextual dependencies at scale (*Vaswani et al., 2017*), and BERT demonstrated the usefulness of deep bidirectional contextual representations for language understanding (*Devlin, Chang, Lee, & Toutanova, 2019*). Foundation models have since expanded the role of large textual datasets in language technology (*Bommasani et al., 2021*). These developments make the linguistic quality of training and evaluation data methodologically important.

If affective language is reduced to isolated labels or emotion nouns, a model may identify happy, sad, angry, or afraid but fail to interpret *What a relief!*, *Seriously?*, or *That made my day*. Corpus-based semantic profiles can provide better evaluation material. A model that distinguishes direct naming, state attribution, reaction formula, evaluative stance, and interpersonal alignment demonstrates stronger pragmatic sensitivity than a model that detects only emotion vocabulary.

BNC and COCA are useful for such evaluation because they provide documented contexts rather than invented examples. BNC can supply British spoken and written evidence, including pre-digital and non-digital interaction. COCA can supply genre-diverse American evidence, including web, blog, subtitle, and spoken-like components. Together, they allow the researcher to test whether a model's interpretation of affective stance changes appropriately across genres, varieties, and contexts.

The limitations of corpora may also be inherited by computational systems. *Bender, Gebru, McMillan-Major, and Shmitchell (2021)* warn that large datasets can reproduce unexamined biases and gaps. In the study of verbal emoticons, this means that underrepresented genres, communities, or multimodal practices may remain underrepresented in model behaviour. Corpus-based research must therefore contribute not only examples but also linguistically justified categories for annotation and evaluation.

8. Conclusions

The article has argued that BNC and COCA are suitable corpus sources for the study of the verbal emoticon when they are used with different expectations. BNC provides documented British evidence, historical comparability, and access to written and spoken material. BNC2014 extends this tradition toward contemporary British usage. COCA provides a large American corpus with genre-sensitive access, diachronic breadth, and tools for frequency, KWIC, collocation, and comparison. Their value lies in grounding interpretation, not in replacing it.

The verbal emoticon should be treated as a functional-discursive category. It is not identical with a graphic emoticon or emoji, and it is not reducible to an isolated emotion word. It includes verbal units and constructions that name, attribute, perform, frame, intensify, or evaluate emotion when they function as markers of affective stance or interpersonal alignment. This definition makes the object corpus-searchable while preserving the need for contextual verification.

The proposed procedure includes inventory building, frequency profiling, concordance reading, collocational and constructional analysis, genre comparison, national-variety comparison, and contextual annotation. The strongest unit of interpretation is the word or construction in a recurrent discourse environment. *Deep fear*, *public anger*, *What a relief!*, *Seriously?*, and *That made my day* are not equivalent objects. They differ in form, explicitness, semantic scope, and pragmatic function.

The study also defined the main limits of corpus interpretation. Frequency measures textual visibility rather than psychological experience. Collocation indicates recurrent co-selection but does not explain meaning without context. Corpus design determines which linguistic evidence can be observed. Multimodal components of digital discourse often remain outside traditional corpus representation. These limits clarify the conditions of philological reliability.

Further research should test the proposed model through systematic extraction of verbal emoticons from BNC and COCA, with normalized frequencies, genre profiles, collocational measures, and annotated concordance samples. A separate continuation should compare corpus-grounded annotation with classifications produced by large language models. Such work would show whether models can recognize not only emotion words but also the subtler verbal mechanisms through which contemporary English encodes affective reaction, evaluative stance, and interpersonal alignment.

References

1. Anokhina, T. O. (2018). *Korpusnyi lakunikon anhliiskoi movy: leksykohrafichnyi analiz* [Corpus lacunicon of the English language: Lexicographic analysis]. *Filolohichni traktaty*, 10(2), 7-12. [https://doi.org/10.21272/Ftrk.2018.10\(2\)-01](https://doi.org/10.21272/Ftrk.2018.10(2)-01) [in Ukrainian].
2. Atkins, S., Clear, J., & Ostler, N. (1992). *Corpus design criteria*. *Literary and Linguistic Computing*, 7(1), 1-16. <https://doi.org/10.1093/lle/7.1.1>

3. Bakhtina, A. O. (2023). *Polilateralnist emoji u kompiuternomu butti (na materiali suchasnykh yevropeyskykh mov)* [The polylaterality of emoji in computer being (based on modern European languages)] (PhD dissertation). Borys Grinchenko Kyiv University, Kyiv. <https://elibrary.kubg.edu.ua/id/eprint/45578/> [in Ukrainian].
4. Bednarek, M. (2008). *Emotion talk across corpora*. London: Palgrave Macmillan. <https://doi.org/10.1057/9780230285712>
5. Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 610-623). <https://doi.org/10.1145/3442188.3445922>
6. Biber, D. (1993). Representativeness in corpus design. *Literary and Linguistic Computing*, 8(4), 243-257. <https://doi.org/10.1093/lc/8.4.243>
7. Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge: Cambridge University Press.
8. BNC Consortium. (2007). *The British National Corpus, version 3 (BNC XML Edition)*. Oxford: Oxford Text Archive.
9. Bober, N. M. (2020). *Matrychne profiluivannia semantyky frazovo-diieslivnykh emotyviv u Brytanskomu natsionalnomu korpusi* [Matrix profiling of the semantics of phrasal-verb emotives in the British National Corpus] (Candidate of Philological Sciences dissertation). National Pedagogical Dragomanov University, Kyiv. https://npu.edu.ua/images/file/vidil_aspirant/dicer/D_26.053.26/Bober1.pdf [in Ukrainian].
10. Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., Bernstein, M. S., Bohg, J., Bosselut, A., Brunskill, E., et al. (2021). On the opportunities and risks of foundation models. *arXiv*. <https://arxiv.org/abs/2108.07258>
11. Brezina, V. (2018). *Statistics in corpus linguistics: A practical guide*. Cambridge: Cambridge University Press.
12. Brezina, V., Hawtin, A., & McEnery, T. (2021). The Written British National Corpus 2014 – design and comparability. *Text & Talk*, 41(5-6), 595-615. <https://doi.org/10.1515/text-2020-0052>
13. British National Corpus Consortium. (2007). *Users reference guide for the British National Corpus XML Edition*. Oxford: Oxford University Computing Services.
14. 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/lc/fqq018>
15. Davies, M. (2020). *The Corpus of Contemporary American English (COCA)*. *English-Corpora.org*. <https://www.english-corpora.org/coca/>
16. Demska-Kulchytska, O. M. (2005). *Osnovy natsionalnoho korpusu ukrainskoi movy* [Foundations of the national corpus of the Ukrainian language]. Kyiv: Instytut ukrainskoi movy NAN Ukrainy [in Ukrainian].
17. Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of NAACL-HLT 2019* (pp. 4171-4186). <https://doi.org/10.18653/v1/N19-1423>
18. Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion*, 6(3-4), 169-200. <https://doi.org/10.1080/02699939208411068>
19. Gries, S. Th. (2017). *Quantitative corpus linguistics with R: A practical introduction* (2nd ed.). New York: Routledge.

20. Gries, S. Th., & Stefanowitsch, A. (2004). Extending collocation analysis: A corpus-based perspective on alternations. *International Journal of Corpus Linguistics*, 9(1), 97-129. <https://doi.org/10.1075/ijcl.9.1.06gri>
21. Holoshchuk, S. L. (2021). Korpusna linhvistyka: suchasnyi stan ta perspektyvy doslidzhen [Corpus linguistics: Current state and prospects of research]. *Current Issues of Linguistics and Translation Studies*, 22, 33-36. <https://doi.org/10.31891/2415-7929-2021-22-7> [in Ukrainian].
22. Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge: Cambridge University Press.
23. Hunston, S. (2007). Semantic prosody revisited. *International Journal of Corpus Linguistics*, 12(2), 249-268. <https://doi.org/10.1075/ijcl.12.2.09hun>
24. Kilgarriff, A. (2001). Comparing corpora. *International Journal of Corpus Linguistics*, 6(1), 97-133. <https://doi.org/10.1075/ijcl.6.1.05kil>
25. Love, R., Dembry, C., Hardie, A., Brezina, V., & McEnery, T. (2017). The Spoken BNC2014: Designing and building a spoken corpus of everyday conversations. *International Journal of Corpus Linguistics*, 22(3), 319-344. <https://doi.org/10.1075/ijcl.22.3.02lov>
26. Makhachashvili, R. (2022). Suchasni vymiry linhvistyky ta komunikatsii: modeli rozvytku movy u tsyfrovomu seredovyschi [Contemporary dimensions of linguistics and communication: Models of language development in the digital environment]. Kyiv: Kyiv Borys Grinchenko University [in Ukrainian].
27. McEnery, T., & Hardie, A. (2012). *Corpus linguistics: Method, theory and practice*. Cambridge: Cambridge University Press.
28. 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>
29. 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>
30. Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.
31. Stefanowitsch, A., & Gries, S. Th. (2003). Collocations: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics*, 8(2), 209-243. <https://doi.org/10.1075/ijcl.8.2.03ste>
32. Tognini-Bonelli, E. (2001). *Corpus linguistics at work*. Amsterdam: John Benjamins. <https://doi.org/10.1075/scl.6>
33. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need. In *Advances in Neural Information Processing Systems*, 30.
34. Wierzbicka, A. (1999). *Emotions across languages and cultures: Diversity and universals*. Cambridge: Cambridge University Press.
35. Zhukovska, V. V. (2023). Kvantyatyvni korpuso-bazovani metody v doslidzhenniakh iz konstruktsiinoi hramatyky [Quantitative corpus-based methods in construction grammar research]. *Visnyk Zhytomyrskoho derzhavnoho universytetu imeni Ivana Franka. Filolohichni nauky*, 1(99), 93-104. [https://doi.org/10.35433/philology.1\(99\).2023.93-104](https://doi.org/10.35433/philology.1(99).2023.93-104) [in Ukrainian].