UDC 811.111'42:004.8
DOI https://doi.org/10.24919/2308-4863/98-1-39

Nataliia BOBER,
orcid.org/0000-0002-9639-0562
Candidate of Philological Sciences, Associate Professor,
Associate Professor at the Department of Germanic Philology
Borys Grinchenko Kyiv Metropolitan University
(Kyiv, Ukraine), n.bober@kubg.edu.ua

THE VERBAL EMOTICON IN ENGLISH-LANGUAGE DIGITAL DISCOURSE AND LARGE LANGUAGE MODELS: A STRUCTURAL-FUNCTIONAL PERSPECTIVE

The article offers an expanded philological interpretation of the verbal emoticon as a heterogeneous set of verbal means through which speakers in contemporary English-language digital discourse do not merely name an emotion but actualise it as an experienced state, an evaluative reaction, a discourse position, or an interpersonal alignment move. In contrast to approaches that reduce verbal emoticons primarily to reactive interjections, the study also includes nominal, adjectival, adverbial, interjective, metaphorical, and syntactically organised constructions whenever they perform an affective and stance-marking function in context. The aim of the article is to determine the structural and functional characteristics of the verbal emoticon in English-language digital discourse and to clarify how these characteristics become visible at the levels of tokenization, embeddings, and syntactic parsing in large language models. The material consists of a micro-corpus of thirty analytical examples reconstructed from recurrent patterns described in recent studies of digital discourse, emotive semantics, interjections, multiword constructions, and LLM-oriented research on subjective language. The methodology combines descriptive, contextual-interpretive, distributive, componential, and hermeneutic-comparative analysis. The study demonstrates that the verbal emoticon should be treated as a macro-category within which emotion may be verbalised through lexical nomination, predicative quality, reactive interjection, graphically intensified form, multiword metaphorical predicate, or exclamative syntactic scheme. It is established that tokenization is most vulnerable when faced with graphically modified, reduced, and phraseological forms; the embedding level preserves the general affective core relatively well but tends to flatten distinctions of intensity, phase structure, and temporal orientation; syntactic parsing is least stable in rendering parenthetical frames of sincerity, nominal exclamatives, fragmentary reactions, and metaphorical unities. The practical value of the article lies in refining philological annotation principles for emotionally marked units, in supporting critical use of LLMs in studies of subjective language, and in expanding digital lexicographic descriptions of verbal means of emotion representation.

Key words: *verbal emoticon, English-language digital discourse, emotional expressivity, tokenization, embeddings, syntactic parsing, large language models.*

Наталія БОБЕР,
orcid.org/0000-0002-9639-0562
кандидат філологічних наук, доцент,
доцент кафедри германської філології
Київського столичного університету імені Бориса Грінченка
(Київ, Україна), n.bober@kubg.edu.ua

ВЕРБАЛЬНИЙ ЕМОТИКОН В АНГЛОМОВНОМУ ЦИФРОВОМУ ДИСКУРСІ ТА ВЕЛИКИХ МОВНИХ МОДЕЛЯХ: СТРУКТУРНО-ФУНКЦІЙНИЙ АСПЕКТ

У статті запропоновано розширене філологічне тлумачення вербального емотикону як сукупності різночастиномовних і різноструктурних вербальних засобів, за допомогою яких у сучасному англомовному цифровому дискурсі мовець не лише називає емоцію, а й переводить її у площину пережитого стану, оцінної реакції, дискурсивної позиції або інтерперсонального налаштування. Дослідження виходить за межі підходів, де вербальний емотикон зводять переважно до реактивних вигуків, і включає субстантивні, ад'єктивні, адвербіальні, вигукові, метафоричні та синтаксично організовані конструкції, якщо в конкретному контексті вони виконують афективно-позиційну функцію. Мета статті полягає у встановленні структурно-функційних характеристик вербального емотикону в англомовному цифровому дискурсі та у з'ясуванні того, як ці характеристики виявляються на токенізаційному, ембедінговому і синтаксично-парсинговому рівнях великих мовних моделей. Матеріалом є мікрокорпус із тридцяти аналітичних прикладів, реконструйованих на підставі повторюваних моделей, описаних у сучасних працях із цифрового дискурсу, емоційної семантики, інтер'єкцій, багатокомпонентних конструкцій і LLM-орієнтованих студіях суб'єктивної мови. Методологія поєднує описовий, контекстуально-

*інтерпретаційний, дистрибутивний, компонентний і герменевтично-зіставний аналіз. Доведено, що **вербальний емотикон** доцільно трактувати як **макрокатегорію**, у межах якої емоція вербалізується через лексичну номінацію, предикативну ознаку, реактивний вигук, графічно інтенсифіковану форму, багатокомпонентний метафоричний предикат або екскламативну синтаксичну схему. Встановлено, що на токенізаційному рівні найуразливішими є графічно модифіковані, редуковані та фразеологізовані форми; на ембедінг-рівні порівняно добре зберігається загальне афективне ядро, але згладжуються відмінності між інтенсивністю, фазовістю та часовою орієнтацією переживання; на синтаксично-парсинговому рівні найменш стабільно репрезентуються вставні рамки щирості, номінативні екскламації, фрагментарні реактиви й метафоричні цілісності. Практичне значення результатів полягає в уточненні принципів філологічної анотації емоційно маркованих одиниць, у критично вивіреному використанні LLM у студіях суб'єктивної мови та в розширенні цифрового лексикографічного опису вербальних засобів емоційної репрезентації.*

Ключові слова: *вербальний емотикон, англомовний цифровий дискурс, емоційна експресивність, токенізація, ембедінг, синтаксичний парсинг, великі мовні моделі.*

Problem Statement. Digital discourse is the linguistic environment in which emotional expressivity most visibly exceeds lexically stabilised naming and gravitates toward flexible, positionally mobile, graphically variable, and syntactically incomplete forms. In such contexts, emotion functions not only as the name of an inner state but also as a way of self-positioning, evaluative response, alignment, distancing, or ironic recalibration of the utterance.

The spread of large language models adds a new dimension to this problem. For the philologist, the crucial question is not merely whether a model recognises positive or negative valence, but which part of the structural organisation of an emotionally marked form is preserved and which part is reduced after tokenization, vectorisation, and structural prediction. The present article, therefore, focuses not on an abstract issue of "emotion in AI" but on a concrete object: the verbal emoticon as a set of verbal means for actualising affective stance in English-language digital discourse.

Analysis of Recent Research and Publications. Studies of digital discourse have convincingly shown that online communication transforms not only the channel of transmission but also the internal organisation of utterances. Hromko treats internet discourse as an innovative linguistic phenomenon defined by multimodality, interactivity, hypertextuality, and anonymity, which means that new norms emerge within actual communicative practice rather than at its margins (Hromko, 2025). Dudko and Zaitseva demonstrate that online language blurs the line between speech and writing and favours forms oriented toward rapid evaluative response, while Svider specifies this tendency on English material by describing the productivity of abbreviations, syntactic simplifications, and hybrid expressive formats (Dudko & Zaitseva, 2024; Svider, 2025).

A second line of work approaches emotional meaning as a dynamic category rather than a fixed dictionary label. Hoemann et al. argue that emotional meaning is constructed in language through attentional framing, appraisal, and the contextual selection of relevant features; their conclusion is essential here because it rules out reducing emotion to a mere lexical tag (Hoemann et al., 2025). Wang and Hu, in their frame-based analysis of knowledge emotion markers, likewise show that affective positioning may be organised as a scenario with a trigger, degree, experiencer, and outcome rather than as an isolated word (Wang & Hu, 2022).

A third group of studies bears directly on the units traditionally associated with verbal emoticons. Dingemanse demonstrates that interjections are not marginal leftovers of language but forms that expose the boundary between reaction, action, and interaction (Dingemanse, 2024). Bober describes verbal emoticons as lexicalised markers of emotional expressivity with a hybrid grammatical status and a tendency toward pragmaticalisation; however, her corpus is centred mainly on reactive units such as haha, ugh, yay, and sigh and therefore leaves unexplored a broader class of verbal means of emotion objectification (Bober, 2025). Honkanen and Müller further show that the choice between verbal and visual reaction is shaped not only by the emotion itself but also by the social parameters of interaction (Honkanen & Müller, 2021).

A separate methodological layer is formed by the works of Rusudan K. Makhachashvili and her co-authors. Emoji Explication in Digital Communication: Logical-Phenomenological Experiment treats emoji as a two-plan digital sign whose meaning emerges at the intersection of expression, content, and interpretation; the conclusion about the polyvalent nature of digital signs is important for the present study because affective meaning in digital communication likewise arises through the intersection of linguistic form and communicative framing (Makhachashvili et al., 2021). Emoji as an Artificial Digital Language shows that interpretation depends on mental frames

and communicative situations, while the studies on AI-enhanced lexicography and data encoding for LLMs emphasise the need for multilayer annotation of innovative units belonging to the digital logosphere (Makhachashvili, Kovpik, & Bakhtina, 2024; Makhachashvili, Semenist, & Klochkov, 2025; Makhachashvili & Bober, 2025).

A fourth block of scholarship addresses the internal representation of language in large language models. Haslett shows that tokenization can alter meaning and is therefore not a neutral preprocessing stage (Haslett, 2025). Wegmann, Nguyen, and Jurgens demonstrate that tokenization is especially sensitive to language variation, including the kinds of orthographic and graphic fluctuation that digital discourse often uses to encode affective intensity (Wegmann et al., 2025). Maheswaran and Desarkar identify relatively stable emotion representations in later layers of LLMs, Kennedy shows that LLMs encode deeper syntactic distinctions than earlier predictive models assumed, and Mititelu et al. emphasise the difficulty of describing multiword expressions in structurally adequate terms (Maheswaran & Desarkar, 2026; Kennedy, 2025; Mititelu et al., 2025). Sanchez-Bayona and Agerri finally demonstrate that metaphor remains an area where models often rely on surface cues more than on integrated interpretive understanding (Sanchez-Bayona & Agerri, 2025).

Taken together, the available scholarship has already described digital discourse as a communicative environment, examined interjections, emotion markers, multiword expressions, and internal emotion representations in LLMs. What remains insufficiently explored is a broad philological account of the verbal emoticon as a system of structurally heterogeneous verbal forms correlated not only with their discourse functions but also with the three levels of model-based representation: tokenization, embeddings, and syntactic parsing. This is the gap the present article seeks to address.

The Aim of the Article. The aim of the article is to determine the structural and functional characteristics of the verbal emoticon in English-language digital discourse and to clarify how these characteristics become visible at the levels of tokenization, embeddings, and syntactic parsing in large language models.

Presentation of the Main Research Material. The study's material is a micro-corpus of 30 analytical examples. These examples are not presented as direct quotations from a single corpus and do not claim to constitute a statistically representative sample; rather, each of them reconstructs recurrent patterns described in studies of digital discourse, interjections, emotive semantics, multiword constructions, and LLM-oriented research on subjective language. This mode of selection enables comparison of structurally distinct forms within a unified analytical framework without conflating heterogeneous platforms, genres, and corpora.

The article combines descriptive, contextual-interpretive, distributive, componential, and hermeneutic-comparative methods. The first describes form and position; the second establishes discourse function in a concrete micro-context; the third traces combinability and syntactic behaviour; the fourth decomposes emotional meaning into valence, intensity, phase structure, and temporal orientation; the fifth compares philologically relevant properties of a form with those aspects that, according to current LLM studies, are retained or weakened at different representational levels.

The verbal emoticon is defined here as a verbal unit or construction which, in a concrete digital utterance, does not simply name an emotion but translates it into the plane of lived experience, evaluation, or stance. This definition deliberately distinguishes the verbal emoticon from neutral metalinguistic naming: fear is one of the basic emotions that describes emotion as a concept, whereas I'm in panic mode after that email turns panic into an experienced, interpreted, and discourse-shaped state.

The material allows five principal structural-functional groups to be distinguished: nominal models, in which emotion is condensed in a nominative centre; predicative-qualifying forms; reactive-interjective units; metaphorical and phraseological constructions; and syntactically organised schemes in which emotion unfolds not in one word but in the architecture of the utterance itself. Structural heterogeneity here is inseparable from functional heterogeneity: nouns condense a state into a compact situation-sign, adjectives and predicatives name an experience directly, interjections transmit primary reaction, metaphorical constructions reconstruct inner experience through an image-scenario, and syntactic frames reveal sincerity, proximity, ironic distance, or exclamative intensity.

Tokenization Level. At the tokenization level, the most sensitive cases are those in which affect is encoded not only by a dictionary word but also by non-standard graphics, colloquial reduction, phraseological indivisibility, or informal framing. It is precisely here that formal segmentation intervenes most directly in meaning.

(1) Soooo relieved we finally got the answer. The affective force of this example is produced not only by the relief but also by the graphically stretched

intensifier, which marks an overflow of post-tension relief. If the form is normalised to neutral, the model preserves positive valence but loses the sign of affective excess (Bober, 2025; Dudko & Zaitseva, 2024; Svider, 2025).

(2) I'm sooo not okay after that thread. The elongated intensifier combines with the colloquial predicative model, not okay, a form that in digital discourse often signals emotional destabilisation without explicit dramatization. Formal segmentation keeps the negativity of the utterance but weakens its staged display of state, where informing and dramatizing coincide (Dudko & Zaitseva, 2024; Svider, 2025).

(3) Ughhh, another "urgent" message at midnight. The elongated interjection verbalises immediate affective rejection, while the quotation marks around the urgent shift the assessment into an ironic mode. At this level, it matters not only that the interjection be isolated as a segment, but that its link to ironic reevaluation of the situation remain intact (Bober, 2025; Dingemanse, 2024).

(4) I'm low-key devastated, actually. The sequence low-key + devastated does not mechanically lower emotional force: low-key functions as a frame of controlled understatement, whereas devastated names a severe affective breakdown. A model that relies only on general negativity smooths out the internal tension between mitigated framing and high-intensity predication (Bober, 2025; Svider, 2025).

(5) Not gonna lie, I'm terrified. The reduced parenthetical frame, not gonna lie, authenticates the utterance and prepares the addressee for an unshielded admission. Wegmann et al.'s point about tokenization sensitivity to language variation is especially relevant here: uneven segmentation threatens not the emotion terrified itself, but the communicative mode in which it is delivered (Bober, 2025; Wegmann et al., 2025).

(6) Gonna lose it if they reschedule again. Emotion is not explicitly named, yet the idiomatic complex 'gonna lose it' signals an approaching affective rupture. The philologically relevant property is the integrity of the expression, not the sum of loss, and it is as separate items; literal segmentation therefore impoverishes the meaning of the whole (Svider, 2025; Wegmann et al., 2025).

(7) Panic mode was activated after that email. The nominal model's panic mode, when activated, turns emotion into an activated behavioural regime. Read word by word, panic remains a noun and mode activated a quasi-technical notification; in digital discourse, however, the whole construction functions as a compact sign of emotional mobilization (Bober, 2025; Dudko & Zaitseva, 2024).

(8) I can't even process this rn. The shortening rn and the discourse operator even jointly encode overload and the temporal immediacy of the response. What matters here is not process as a neutral predicate but the inability to process as a symptom of strong affect; once the temporal marker is weakened, the state is detached from the moment of utterance (Svider, 2025; Dudko & Zaitseva, 2024; Wegmann et al., 2025).

(9) My heart sank when his name popped up. This is not a free word combination but a conventional metaphorical predicate of sudden emotional decline. Technically correct segmentation is not yet a philologically adequate interpretation if the model fails to preserve the figurative integrity of the expression (Mititelu et al., 2025; Sanchez-Bayona & Agerri, 2025).

(10) This comment hit a nerve, badly. The emotion is verbalised through a phraseological predicate without an explicit emotion word, but with a clear scenario of painful affective contact. Tokenization that splits the phrase into literal hit and nerve keeps only part of the signal and loses the emotionally relevant indivisibility of the construction (Mititelu et al., 2025; Sanchez-Bayona & Agerri, 2025).

These examples show that tokenization is most fragile where emotional meaning is distributed across non-standard graphics, colloquial framing, and idiomatic indivisibility. In such cases, formal segmentation may be technically correct and yet philologically insufficient, because what is reduced is not an isolated word but a mode of experiencing encoded in the shape of the utterance.

Embeddings Level. At the embeddings level, the issue is no longer how a form is segmented, but which parameters of emotional meaning the model retains as shared and which it levels out. Here it becomes especially clear that neighbourhood in semantic space does not amount to identity of structural-functional profile.

(11) The update brought real relief at last. Relief belongs to the positive field, but its semantic core includes prior tension and its release. A model that registers only positivity loses the emotion's retrospective structure – the fact that relief presupposes anxiety, waiting, or uncertainty as its precursor (Hoemann et al., 2025; Wang & Hu, 2022).

(12) Her post radiates joy, not just approval. Joy is explicitly contrasted with the more neutral approval, suggesting that the utterance presents emotion as a lived state rather than a simple positive judgment. Proximity within a positive semantic field, therefore, does not explain the difference between delight, approval, and joy as experienced affect (Hoemann et al., 2025).

(13) I'm honestly devastated by the result. Devastated marks not situational irritation but a state of emotional desolation following an event of high personal significance. Models usually align devastated with other strongly negative items, yet they often flatten the depth and consequentiality of their post-event experiences (Maheswaran & Desarkar, 2026; Hoemann et al., 2025).

(14) I'm ridiculously anxious about tomorrow. Anxious belongs to prospective emotional states: its core is tense anticipation of a future event. If embeddings collapse anxious, terrified, and devastated into a single negative field, they erase the distinction between anticipation, peak fear, and post-factum emotional ruin (Maheswaran & Desarkar, 2026).

(15) He sounded oddly cheerful after the breakup. Cheerful does not function here as a simple positive quality but as a marker of a contextually anomalous reaction. The adverb 'oddly' makes positivity interpretively unstable, and it is precisely this tension that the generalised vector neighbourhood tends to smooth out (Hoemann et al., 2025; Wang & Hu, 2022).

(16) I'm on edge waiting for the reply. The phraseological construction on edge combines a bodily metaphor of unstable balance with prospective anxiety. Models usually identify their negative affective core, but they preserve their specific profile less reliably, remaining in a state of tense waiting without transitioning to full panic (Mititelu et al., 2025; Maheswaran & Desarkar, 2026).

(17) I'm heartbroken, not merely sad. The contrast between heartbreak and sadness explicitly foregrounds the distinction between deep personal loss and ordinary sadness. In semantic space, these items will naturally cluster, yet for philological analysis, heartbroken includes the idea of inner rupture rather than simple negative valence (Sanchez-Bayona & Agerri, 2025).

(18) What a nightmare of a day. Nightmare in this position does not simply denote a dream or even fear; it evaluates the entire event as an emotionally exhausting whole. Embeddings capture the negative core of the word, but they do not reliably preserve its exclamatory-evaluative force or its connection to a full scenario of intolerable experience (Hoemann et al., 2025; Sanchez-Bayona & Agerri, 2025).

(19) Wow, what a relief. The combination of wow and relief joins immediate reaction with secondary interpretation of the state. Models tend to align relief with positive affect, yet they do not always preserve the two-stage organisation of the experience: first, the sudden break in tension, then its explicit verbal recognition (Dingemanse, 2024; Maheswaran & Desarkar, 2026).

(20) Damn, that hurt. In this construction, hurt cannot be read simply as a neutral description of pain; together with damn, it signals offence, emotional injury, and abrupt affective reaction. Vector proximity to the vocabulary of pain is possible, but it does not exhaust the pragmatic force of this short digital utterance (Dingemanse, 2024; Honkanen & Müller, 2021).

The material in this section demonstrates that embeddings confidently gather emotionally related units into common fields, but for exactly that reason, they also smooth out what matters most to philological analysis: the distinction between prospective anxiety and post-event desolation, between joy and approval, between sadness and felt rupture. Vector proximity does not cancel structural-functional heterogeneity.

Syntactic-Parsing Level. The syntactic-parsing level reveals not only where an emotional lexeme is located in the sentence, but how emotion is built into the communicative architecture of the utterance. Here, it becomes particularly clear that the philologically relevant unit is often wider than a single predicate or a single noun.

(21) Honestly, I'm devastated by the delay. The parenthetical honestly does not duplicate the content of devastated; it frames the predication as sincere self-disclosure. Parsing that reduces this frame to an optional adverb weakens the communicative mode in which the emotion is addressed to the interlocutor (Kennedy, 2025).

(22) To my surprise, I felt real relief after reading it. The initial frame, to my surprise, establishes an interpretive regime before relief appears: the emotion is presented as unexpected in relation to prior expectation. What matters syntactically is therefore not only the main predicate but also the frame that alters its affective configuration (Kennedy, 2025; Wang & Hu, 2022).

(23) What a relief that was. Emotion is located not only in relief as a noun but in the full exclamative scheme, what a + noun. If the model reads relief simply as a positive lexeme, it misses the syntactic mechanism of intensification that turns the word into an affective reaction rather than a neutral nomination (Kennedy, 2025).

(24) What a nightmare this became. Nightmare functions as the nominal centre of an exclamatory evaluation, shifting the whole utterance into a dynamic plane: the event has become an unbearable experience. Without the scheme as a whole, parsing impoverishes the utterance into a negative noun plus an auxiliary clause (Kennedy, 2025; Sanchez-Bayona & Agerri, 2025).

(25) Ugh, another last-minute change. Formally, this is an incomplete utterance, yet in digital discourse,

it is a fully sufficient emotional reply. The interjection sets the reaction, and the nominal fragment names its trigger; syntactic incompleteness is therefore not a defect but a condensed form of communicative organisation (Dingemanse, 2024; Bober, 2025).

(26) Wow, you actually made it! Wow serves as an immediate reaction, while actually in the following clause marks previously lowered expectations. The emotional effect emerges from the sequence "reaction – correction of expectation – interpersonal convergence", so the interaction between interjection and clause is crucial (Dingemanse, 2024; Honkanen & Müller, 2021).

(27) If I'm being honest, I'm still anxious. The clause, if I'm being honest, introduces not a condition but a frame of self-disclosure that strengthens the credibility of anxious. The adverb still adds duration, and it is precisely the combination of the sincerity frame and the durative marker that creates a compact structure of a sustained, consciously registered experience (Kennedy, 2025; Hoemann et al., 2025).

(28) It broke me, if only for a minute. The metaphorical predicate broke me expresses not literal breakage but a momentary affective collapse. The attached frame is only for a minute, moderating intensity without cancelling it; parsing must therefore preserve the proportion between maximal emotional wording and minimising temporal qualification (Sanchez-Bayona & Agerri, 2025; Kennedy, 2025).

(29) The post sounded cheerful, but the replies were furious. The conjunction but separates two emotional perspectives within one discourse fragment: the initial message and the collective reaction to it. This matters in digital environments, where affect is often distributed between author and audience rather than localised in a single speaker (Kennedy, 2025; Hromko, 2025).

(30) I'm in panic mode because the thread will not stop. The subordinate causal clause explains not panic as an abstract lexical fact but its discourse trigger. The syntactic structure reveals the causal infrastructure of the experience: emotion is presented as a response to the continuity of digital stimulus rather than as a self-contained inner state (Bober, 2025; Kennedy, 2025).

A comparison of the three levels of representation makes it possible to specify that large language models preserve the affective core of an expression more steadily than those properties that are crucial for philological description: graphic intensity, metaphorical integrity, sincerity framing, temporal orientation, and the syntactic scheme of reaction. Put differently, the model can often identify the emotional field, but it reproduces the way that field is built by concrete linguistic organisation less reliably.

For this reason, tokenization, embeddings, and parsing cannot be treated as merely technical stages. In emotionally marked digital discourse, they are directly related to whether a model preserves graphic amplification, idiomatic indivisibility, phase structure of emotion, distribution of affective roles between interactants, or the exclamative completeness of a formally incomplete utterance. Philological interpretation is therefore not an external comment on the algorithm but a condition of adequate object description.

Conclusions. The analysis supports a broad understanding of the verbal emoticon in contemporary English-language digital discourse as a macro-category of structurally heterogeneous verbal means that actualise the speaker's affective position in context. At the tokenization level, graphically modified, reduced, and phraseological forms prove most vulnerable; at the embeddings level, differences in intensity, phase structure, and temporal orientation tend to be flattened; at the syntactic-parsing level, parenthetical frames of sincerity, exclamative schemes, fragmentary reactions, and multiword metaphorical unities are reproduced least consistently. The practical significance of the results lies in refining annotation principles for emotionally marked units, enabling more critical use of LLMs in studies of subjective language, and improving digital lexicographic description of verbal means of emotion representation.

BIBLIOGRAPHY

1. Bober, N. M. Verbal Emoticons as Markers of Emotional Expressivity: A Corpus-Linguistic Analysis of English Digital Discourse. *Scientific Journal of Mykhailo Dragomanov State University of Ukraine*. Series 9. Current Trends in Language Development. 2025. Vol. 30. P. 16–25. DOI: https://doi.org/10.31392/UDU-nc.series9.2025.30.02

2. Dingemanse, M. Interjections at the Heart of Language. *Annual Review of Linguistics.* 2024. Vol. 10. P. 257–277. DOI: https://doi.org/10.1146/annurev-linguistics-031422-124743

3. Dudko, I., Zaitseva, N. Modern Discourse of Internet Communication: Linguistic Aspect. *International Journal of Philology*. 2024. Vol. 15(3). P. 9–21. DOI: https://doi.org/10.31548/philolog/3.2024.09

4. Haslett, D. A. Tokenization Changes Meaning in Large Language Models: Evidence from Chinese. *Computational Linguistics*. 2025. Vol. 51(3). P. 785–814. DOI: https://doi.org/10.1162/coli_a_00557

5. Hoemann, K., Lee, Y., Dussault, É., Devylder, S., Ungar, L. H., Geeraerts, D., Mesquita, B. The Construction of Emotional Meaning in Language. *Communications Psychology.* 2025. Vol. 3. Article 99. DOI: https://doi.org/10.1038/s44271-025-00255-0

6. Honkanen, M., Müller, J. Interjections and Emojis in Nigerian Online Communication. *World Englishes.* 2021. Vol. 40(4). P. 611–630. DOI: https://doi.org/10.1111/weng.12544

7. Hromko, T. V. Theoretical Dimension of Internet Discourse as an Innovative Linguistic Phenomenon. Scientific Notes of V. I. *Vernadsky Taurida National University.* Series: Philology. Journalism. 2025. Vol. 36(75), Part 1. P. 244–254. DOI: https://doi.org/10.32782/2710-4656/2025.1.1/40

8. Kennedy, M. Evidence of Generative Syntax in LLMs. In: *Proceedings of the 29th Conference on Computational Natural Language Learning*. 2025. P. 377–396. DOI: https://doi.org/10.18653/v1/2025.conll-1.25

9. Maheswaran, A., Desarkar, M. S. A Unified View on Emotion Representation in Large Language Models. In: *Proceedings of the 19th Conference of the European Chapter of the Association for Computational Linguistics*. 2026. P. 3589–3610. URL: https://aclanthology.org/2026.eacl-long.165/

10. Makhachashvili, R., Bakhtina, A., Semenist, I., Prihodko, G., Prykhodchenko, O. Emoji Explication in Digital Communication: Logical-Phenomenological Experiment. In: *Digital Humanities Workshop* (DHW 2021). ACM, New York, NY, 2021. DOI: https://doi.org/10.1145/3526242.3526259

11. Makhachashvili, R. K., Kovpik, S. I., Bakhtina, A. O. Emoji as an Artificial Digital Language: A Frame-Semantic Analysis of Perception and Interpretation. *CEUR Workshop Proceedings*. 2024. Vol. 3820. P. 146–163.

12. Makhachashvili, R., Bober, N. AI-Enhanced Transdisciplinary Data Encoding for LLMs Training. *Proceedings of the 29th World Multi-Conference on Systemics, Cybernetics and Informatics* (WMSCI 2025). 2025. P. 327–333. DOI: https://doi.org/10.54808/WMSCI2025.01.327

13. Makhachashvili, R., Semenist, I., Klochkov, V. AI-Enhanced Multilingual Lexicography for Digital Communication. *Proceedings of the 16th International Multi-Conference on Complexity, Informatics and Cybernetics* (IMCIC 2025). 2025. P. 247–253. DOI: https://doi.org/10.54808/IMCIC2025.01.247

14. Mititelu, V. B., Giouli, V., Korvel, G., Liebeskind, C., Lobzhanidze, I., Makhachashvili, R., Markantonatou, S., Markovic, A., Stoyanova, I. The Challenges of Syntactic Descriptions of Multiword Expressions in Electronic Lexicography. In: *Electronic Lexicography in the 21st Century* (eLex 2025): Intelligent Lexicography. 2025. P. 1–20.

15. Sanchez-Bayona, E., Agerri, R. Metaphor and Large Language Models: When Surface Features Matter More than Deep Understanding. In: *Findings of the Association for Computational Linguistics*: ACL 2025. 2025. P. 17462–17477. DOI: https://doi.org/10.18653/v1/2025.findings-acl.898

16. Svider, I. A. Linguistic Features of Digital English Discourse. *Philological Periodical of Lviv.* 2025. №. 17. P. 109–114. DOI: https://doi.org/10.32447/2663-340X-2025-17.15

17. Wang, Q., Hu, G. What Surprises, Interests and Confuses Researchers? A Frame-Based Analysis of Knowledge Emotion Markers in Research Articles. *Lingua*. 2022. Vol. 279. Article 103426. DOI: https://doi.org/10.1016/j.lingua.2022.103426

18. Wegmann, A., Nguyen, D., Jurgens, D. Tokenization Is Sensitive to Language Variation. In: *Findings of the Association for Computational Linguistics*: ACL 2025. 2025. P. 10958–10983. DOI: https://doi.org/10.18653/v1/2025.findings-acl.572

REFERENCES

1. Bober, N. M. (2025). Verbal emoticons as markers of emotional expressivity: A corpus-linguistic analysis of English digital discourse. Scientific Journal of Mykhailo Dragomanov State University of Ukraine. Series 9. Current Trends in Language Development, 30, 16–25. https://doi.org/10.31392/UDU-nc.series9.2025.30.02

2. Dingemanse, M. (2024). Interjections at the heart of language. Annual Review of Linguistics, 10, 257–277. https://doi.org/10.1146/annurev-linguistics-031422-124743

3. Dudko, I., & Zaitseva, N. (2024). Modern discourse of internet communication: Linguistic aspect. International Journal of Philology, 15(3), 9–21. https://doi.org/10.31548/philolog/3.2024.09

4. Haslett, D. A. (2025). Tokenization changes meaning in large language models: Evidence from Chinese. Computational Linguistics, 51(3), 785–814. https://doi.org/10.1162/coli_a_00557

5. Hoemann, K., Lee, Y., Dussault, É., Devylder, S., Ungar, L. H., Geeraerts, D., & Mesquita, B. (2025). The construction of emotional meaning in language. Communications Psychology, 3, Article 99. https://doi.org/10.1038/s44271-025-00255-0

6. Honkanen, M., & Müller, J. (2021). Interjections and emojis in Nigerian online communication. World Englishes, 40(4), 611–630. https://doi.org/10.1111/weng.12544

7. Hromko, T. V. (2025). Theoretical dimension of internet discourse as an innovative linguistic phenomenon. Scientific Notes of V. I. Vernadsky Taurida National University. Series: Philology. Journalism, 36(75), Part 1, 244–254. https://doi.org/10.32782/2710-4656/2025.1.1/40

8. Kennedy, M. (2025). Evidence of generative syntax in LLMs. In Proceedings of the 29th Conference on Computational Natural Language Learning (pp. 377–396). https://doi.org/10.18653/v1/2025.conll-1.25

9. Maheswaran, A., & Desarkar, M. S. (2026). A unified view on emotion representation in large language models. In Proceedings of the 19th Conference of the European Chapter of the Association for Computational Linguistics (pp. 3589–3610).

10. Makhachashvili, R., Bakhtina, A., Semenist, I., Prihodko, G., & Prykhodchenko, O. (2021). Emoji explication in digital communication: Logical-phenomenological experiment. In Digital Humanities Workshop (DHW 2021). ACM. https://doi.org/10.1145/3526242.3526259

11. Makhachashvili, R., & Bober, N. (2025). AI-enhanced transdisciplinary data encoding for LLMs training. In Proceedings of the 29th World Multi-Conference on Systemics, Cybernetics and Informatics (WMSCI 2025) (pp. 327–333). https://doi.org/10.54808/WMSCI2025.01.327

12. Makhachashvili, R. K., Kovpik, S. I., & Bakhtina, A. O. (2024). Emoji as an artificial digital language: A frame-semantic analysis of perception and interpretation. CEUR Workshop Proceedings, 3820, 146–163.

13. Makhachashvili, R., Semenist, I., & Klochkov, V. (2025). AI-enhanced multilingual lexicography for digital communication. In Proceedings of the 16th International Multi-Conference on Complexity, Informatics and Cybernetics (IMCIC 2025) (pp. 247–253). https://doi.org/10.54808/IMCIC2025.01.247

14. Mititelu, V. B., Giouli, V., Korvel, G., Liebeskind, C., Lobzhanidze, I., Makhachashvili, R., Markantonatou, S., Markovic, A., & Stoyanova, I. (2025). The challenges of syntactic descriptions of multiword expressions in electronic lexicography. In Electronic Lexicography in the 21st Century (eLex 2025): Intelligent lexicography (pp. 1–20).

15. Sanchez-Bayona, E., & Agerri, R. (2025). Metaphor and large language models: When surface features matter more than deep understanding. In Findings of the Association for Computational Linguistics: ACL 2025 (pp. 17462–17477). https://doi.org/10.18653/v1/2025.findings-acl.898

16. Svider, I. A. (2025). Linguistic features of digital English discourse. Philological Periodical of Lviv, 17, 109–114. https://doi.org/10.32447/2663-340X-2025-17.15

17. Wang, Q., & Hu, G. (2022). What surprises, interests and confuses researchers? A frame-based analysis of knowledge emotion markers in research articles. Lingua, 279, Article 103426. https://doi.org/10.1016/j.lingua.2022.103426

18. Wegmann, A., Nguyen, D., & Jurgens, D. (2025). Tokenization is sensitive to language variation. In Findings of the Association for Computational Linguistics: ACL 2025 (pp. 10958–10983). https://doi.org/10.18653/v1/2025.findings-acl.572

Дата першого надходження статті до видання: 15.04.2026
Дата прийняття статті до друку після рецензування: 05.05.2026
Дата публікації (оприлюднення) статті: 25.05.2026