Bober N. M.
Candidate of Philological Sciences, Associate Professor,
Associate Professor at the Department of Germanic Philology,
Faculty of Romance and Germanic Philology,
(*Borys Grinchenko Kyiv Metropolitan University*)

Verbal emoticon in cognitive-emotive NLP modelling of digital discourse

The rapid development of artificial intelligence has changed the way foreign languages are taught, learned, and described philologically. In digital communication, emotional meaning is increasingly transmitted not only through direct emotional words but also through compressed reactive constructions that imitate facial expressions, gestures, laughter, shock, or affective overload. This paper focuses on the verbal emoticon as one of such units. In the proposed interpretation, a verbal emoticon is a lexical or phraseological digital marker that performs the communicative function of an emoji, emoticon or paralinguistic reaction, while remaining formally verbal: *I'm dead, I'm crying, I can't, screaming, shaking, ugh, wow*.

The relevance of this issue lies in a methodological gap between philological interpretation and automatic text processing. Emotion analysis in NLP is a rapidly developing field, but its terminology, emotion taxonomies and annotation principles remain heterogeneous (Plaza-del-Arco et al., 2024). Traditional sentiment analysis usually reduces evaluation to positive, negative or neutral polarity. However, verbal emoticons show that the emotional meaning of a digital utterance cannot be derived solely from polarity. The expression *I'm dead* may express amusement, surprise or admiration rather than death; *I'm crying* may indicate sadness, laughter, tenderness or ironic affect, depending on the context.

The aim of these theses is to outline a cognitive-emotive philological algorithm for NLP-based analysis of verbal emoticons as emotionally marked linguistic units. The theoretical assumption is that a verbal emoticon should not be treated as an isolated token with a fixed dictionary meaning. It should be analysed as a contextual sign in which lexical form, evaluation, intensity and pragmatic function interact. This position aligns with componential approaches to emotion, which view emotion as connected to cognitive appraisal and situational evaluation rather than a single lexical label (Scherer, 2005).

The proposed algorithm contains seven reproducible stages. First, the researcher forms a corpus of English-language digital texts: social media comments, short reactions, educational chat fragments, online reviews or AI-mediated dialogue samples. Second, linguistic preprocessing is performed: segmentation, tokenization, lemmatization and part-of-speech identification. Negation, intensifiers and discourse markers must be preserved, because not, never, so, literally, just and very may radically change emotional interpretation.

The third stage is the identification of emotional markers. A preliminary lexical search may rely on emotion lexicons, such as the NRC Emotion Lexicon, which links English words to basic emotions and polarity (Mohammad & Turney, 2013). Nevertheless, the lexicon should be used only as an auxiliary instrument. It detects direct emotion words more reliably than non-literal digital reactions. Therefore, the annotation must include not only direct nominations such as *fear, anger, joy* or *sadness*, but also evaluative adjectives, emotional verbs, interjections and verbal emoticons.

The fourth stage is double annotation. Each unit is marked at two levels: surface emotionality and contextual emotionality. Surface emotionality records what is explicitly present in the text. Contextual emotionality records the emotion inferred from the communicative situation. For example, "*I am angry*" is an explicit nomination, whereas "*My hands were shaking when I opened the message*" signals an implicit emotional state. Verbal emoticons usually belong to the borderline zone: they are formally visible in the utterance, but their emotional meaning is often indirect, hyperbolic or ironic.

The fifth stage is the cognitive-evaluative matrix. Each example is described according to the following parameters: emotional category, explicit or implicit expression, evaluative polarity, intensity, source of emotion, emotional subject, type of linguistic marker and pragmatic function. The utterance *I’m dead after reading this* may therefore be annotated as a verbal emoticon with high intensity, non-literal meaning and the function of hyperbolised reaction. Depending on the broader context, its probable emotional categories may be amusement, surprise, admiration or shock.

The sixth stage is contextual modelling. Transformer-based models are relevant because they create contextual representations rather than process words as isolated units. BERT was designed to pre-train bidirectional representations by jointly conditioning on left- and right-context (Devlin et al., 2019). This is crucial for verbal emoticons: *dead* in a *dead body* and *dead* in *I’m dead*, this is hilarious, requires different computational representations. GoEmotions also shows that contemporary NLP increasingly moves beyond simple polarity and models a wider range of emotion categories (Demszky et al., 2020).

The seventh stage is philological verification of machine output. The model’s classification is compared with manual annotation. Typical errors include literalizing figurative expressions, confusing laughter and sadness, ignoring negation, weak recognition of irony, and underestimating intensity. These errors reveal where machine interpretation lacks pragmatic knowledge. Emoji-based distant supervision has shown that digital emotional signs can improve representations for sentiment, emotion and sarcasm detection (Felbo et al., 2017). Verbal emoticons deserve similar attention because they are verbal, teachable and suitable for annotation.

The proposed approach is also applicable to foreign language education. Learners who communicate in English in AI-mediated environments encounter digital emotional patterns that are absent from many traditional textbooks. Teaching verbal emoticons as contextual emotional markers helps students distinguish literal meaning from pragmatic function. At the same time, the algorithm trains critical work with AI tools: students may compare human interpretation with automatic emotion recognition and identify cases where the model simplifies, misclassifies or decontextualizes emotional meaning.

Thus, the verbal emoticon should be regarded as a complex unit of digital discourse, not as a marginal colloquial form. It combines nomination, evaluation, expressiveness and pragmatic action. Its NLP modelling requires double annotation, cognitive-evaluative classification, contextual representation and philological verification. Such a methodology may improve emotion analysis in NLP and provide a practical bridge between artificial intelligence, corpus linguistics and foreign language education.

References:

1. Demszky, D., Movshovitz-Attias, D., Ko, J., Cowen, A., Nemade, G., & Ravi, S. (2020). GoEmotions: A dataset of fine-grained emotions. ACL 2020, 4040–4054. https://doi.org/10.18653/v1/2020.acl-main.372

2. Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. NAACL-HLT 2019, 4171–4186. https://doi.org/10.18653/v1/N19-1423

3. Felbo, B., Mislove, A., Søgaard, A., Rahwan, I., & Lehmann, S. (2017). Using millions of emoji occurrences to learn any-domain representations for detecting sentiment, emotion and sarcasm. EMNLP 2017, 1615–1625. https://doi.org/10.18653/v1/D17-1169

4. 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

5. Plaza-del-Arco, F. M., Cercas Curry, A. A., Cercas Curry, A., & Hovy, D. (2024). Emotion analysis in NLP: Trends, gaps and roadmap for future directions. LREC-COLING 2024, 5696–5710.

6. 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