

## Assessing emotion in L2 writing: Validating Watson NLU with emotional vocabulary training

### *Evaluación de las emociones en la escritura en L2: Validación de Watson NLU con entrenamiento de vocabulario emocional*

María Jesús Sánchez <sup>a</sup>, Elisa Pérez-García <sup>b</sup>, Beatriz Bermúdez-Margaretto<sup>c</sup>

<sup>a</sup> Universidad de Salamanca, [mjs@usal.es](mailto:mjs@usal.es)

<sup>b</sup> Universidad de Salamanca, [elisapg@usal.es](mailto:elisapg@usal.es)

<sup>c</sup> Universidad de Salamanca, [bermudezmargaretto@usal.es](mailto:bermudezmargaretto@usal.es)

#### Abstract

Affective word values have been widely studied across languages, often focusing on isolated words due to the difficulty of assessing emotionality in texts. This study examines whether written emotional content can be reliably captured using a specific software tool (Watson Natural Language Understanding). Thirty-three Spanish undergraduates wrote 150-word autobiographical texts in their L2 (English) before and after training with emotional vocabulary. Normative valence ratings of content words obtained in the pre- and post-training phases were compared with sentiment scores generated by Watson NLU. Strong positive relations were found between sentiment and normative valence scores in both phases, with stronger relations at post-training. Regression analyses confirmed that sentiment scores significantly predicted normative valence. Importantly, while normative valence did not differ between phases, sentiment scores increased after training. These results suggest that Watson NLU is a valid and sensitive tool for assessing emotionality in written language and its modulation through training during text writing.

Keywords. Emotion, bilingualism, sentiment, valence, emotional training

#### Resumen

*Los valores afectivos de las palabras se han estudiado ampliamente en distintas lenguas, a menudo centrándose en palabras aisladas debido a la dificultad de evaluar la emocionalidad en los textos. Este estudio analiza si el contenido emocional escrito puede captarse de forma fiable mediante Watson Natural Language Understanding. Treinta y tres universitarios españoles escribieron textos autobiográficos de 150 palabras en su L2 (inglés) antes y después de un entrenamiento en vocabulario emocional. La valencia normativa de las palabras de contenido se comparó con las puntuaciones de sentimiento generadas por Watson NLU. Ambas medidas mostraron correlaciones positivas y fuertes en las fases pre- y post-entrenamiento, siendo mayores tras el entrenamiento. Los análisis de regresión confirmaron que las puntuaciones de sentimiento predijeron significativamente la valencia normativa. Aunque no se observaron cambios en la valencia normativa, las puntuaciones de sentimiento aumentaron tras el entrenamiento, lo que indica la sensibilidad de la herramienta a la modulación emocional del lenguaje durante la escritura de textos.*

*Palabras clave. Emoción, bilingüismo, tono emocional, valencia, entrenamiento emocional*

DOI: 10.26378/rnlael2040659

Recibido: 09/01/2026 - Aprobado: 1/04/2026

Publicado bajo licencia de Creative Commons Reconocimiento Sin Obra Derivada 4.0 Internacional

### 1. Introduction

Word emotionality has been investigated in first (L1) and second (L2) languages (Imbault et al., 2021; Warriner et al., 2013) by means of different measures: subjective ratings (Dewaele, 2004; Pavlenko, 2005), more objective, behavioral data (response times or accuracy rates during reading or word categorization), or physiological (skin conductance, Harris, 2004; Harris et al., 2006) and neurophysiological indices (neural responses, Opitz & Degner, 2012). This research has shown that emotional words are perceived and evaluated as more emotionally extreme (i.e., more positive or negative) in L1 than in L2 (Caldwell-Harris, 2015; Ferré et al., 2010, Sánchez et al., 2025) and that their processing is more automatic and effortful in the L1 too, leading to higher physiological reactivity (skin conductance, electromyography, pupillometry) and increased or faster brain responses (Conrad et al., 2011; Fan et al., 2018; Foroni, 2015; Opitz & Degner, 2012; Toivo & Scheepers, 2019; Winskel, 2013). Although results in the production domain are scarcer, recent evidence also shows higher emotional verbal fluency in L1 than L2 (Lam & Mardquardt, 2022) as well as more diverse emotional vocabulary during L1 than L2 text production (Pavlenko & Driagina, 2007, Kyriakou et al., 2024; Vidal Noguera & Mavrou, 2025). Similarly, gestures usage has been reported during the retelling of emotional experiences in L1 than L2 (Emir Özder et al., 2023). Overall, this research systematically highlights that L2 speakers tend to experience their L2 as less emotional.

Nonetheless, there is still no unanimous conclusion on the reduced emotional sensitivity in L2 (see, for instance, lack of L1-L2 differences in Eilola & Havelka, 2010; Kazanas & Altarriba, 2016), with various factors such as age of L2 acquisition, proficiency and exposure potentially modulating the L2 emotionality (Conrad et al., 2011; Opitz & Degner, 2011). In this line of research, the common reduced emotionality in L2 has been attributed to weaker associations between words and their emotional contexts, largely due to fewer meaningful encounters with those words (Pavlenko, 2012). Unlike L1, L2 is often learned in formal instructional contexts such as the school or university, where language use tends to be less spontaneous and less embedded in emotionally rich interactions. As a result, the limited exposure to words in socially and affectively meaningful contexts has been proposed as a key factor underlying the lower emotional resonance in L2.

From this perspective, it is reasonable to hypothesize that increasing the number of such encounters and embedding learning in richer emotional contexts—as implemented in the present study—may help counteract this reduced emotionality. In this sense, this study aims to examine whether a pedagogical intervention enhances the emotional content of learners' written production by increasing meaningful encounters with emotional vocabulary and promoting deeper lexical-semantic processing. Among the instructional strategies that may help enhance emotionality in L2 within formal learning contexts, the summary strategy appears particularly relevant. Summarizing involves reducing a source text to its essential ideas and therefore requires a demanding higher-order cognitive process in which learners synthesize content and identify the most relevant information (Khoshsima & Rabani Nia, 2014). This complex task engages both cognitive and metacognitive operations, including scanning,

skimming, inferencing, and information construction (Keck, 2006; Mokeddem & Houcine, 2016), making reading and writing closely interdependent. Previous research has shown that the use of summarization promotes reading comprehension, writing development, and vocabulary acquisition (Keck, 2014; Shokrpour et al., 2013; Stevens et al., 2019; Hsiang et al., 2020), likely because it encourages deeper lexical-semantic processing. Such effortful processing may strengthen the mental representation of L2 words and facilitate vocabulary learning, including emotionally charged lexical items. Thus, it was expected that producing personal summaries of texts containing both positive and negative emotional content would increase learners' encounters with emotional vocabulary in meaningful contexts, and that this repeated, elaborative engagement would promote a stronger integration of emotional language in L2.

When assessing emotional content in written texts, two complementary approaches can be adopted: focusing on the emotional value of individual words or analysing emotion at the level of the text as a whole. Previous studies on emotional processing in L2 have predominantly examined isolated words (Kousta et al., 2009; Opitz & Degner, 2012; Palazova et al., 2011), whereas evidence at sentence or text level remains comparatively limited (Tang & Ding, 2024; Sheikh & Titone, 2016; Vidal Noguera & Mavrou, 2025; Kyriakou et al., 2024), largely because capturing the emotional tone of an entire text is methodologically more complex. To address this limitation, the present study also explores whether emotional content in written production in L2 can be captured using recently developed AI-based tools, specifically IBM Watson Natural Language Understanding (hereafter, Watson NLU<sup>1</sup>). This natural language processing system extracts meaning from both structured and unstructured language data and can provide information about sentiment expressed in text. Unlike traditional approaches based solely on lexical items, this tool estimates sentiment at the phrase or text level, assigning a score on a continuum from negative to positive (from -1 to +1), thereby allowing the analysis of the overall emotional tone conveyed in a text rather than only the valence of isolated words.

At this point, it is worth clarifying the distinction between lexical valence analysis and sentiment analysis carried out at single-word or text level, respectively. Lexical valence refers to the affective polarity associated with individual words along a bipolar continuum from negative to positive, traditionally derived from normative affective databases such as those developed by Warriner et al. (2013) in L1 English. These databases provide emotional ratings for isolated words based on native-speaker judgments, while comparable L2 norms for bilinguals and foreign language learners remain scarce (Imbault et al., 2021). By contrast, sentiment refers to the overall evaluative attitude or emotional tone expressed by the writer toward a topic within discourse, making it inherently context-sensitive. Although the terms emotion and sentiment are sometimes used interchangeably, emotion typically refers to internal affective states, whereas sentiment reflects how those states are linguistically conveyed in context. From this perspective, sentiment analysis may offer a more naturalistic way of assessing emotional tone in writing, because it evaluates meaning at text level rather than assigning normative polarity to lexical items (Taherdoost & Madanchian, 2023). In addition, AI-based tools like Watson NLU provide important methodological advantages: they enable consistent, replicable analysis of large text samples and reduce

the degree of subjectivity associated with manual coding procedures (Pérez-García & Sánchez, 2020).

Regarding Watson-based sentiment tools, the earlier version of Watson NLU, IBM Watson Tone Analyzer, has been shown to be particularly useful for examining emotional features of language in diverse contexts. For example, Maleki et al. (2023) investigated whether financial incentives influence the production of health-related content on social media by comparing posts from Steemit, a platform that rewards user participation, with posts from Reddit, where no such incentives are provided. Their analysis showed that posts written in the incentive-based environment displayed a more confident and analytical language style, were less tentative, and expressed more joy and less negativity than those published on the non-incentivized platform. Similarly, Steffens et al. (2021) used Watson Tone Analyzer to examine whether the source of funding influences how medical research findings are written. By examining emotional features in the texts—such as expressions of anger, fear, joy, and sadness—as well as language style (for example, whether the writing sounded more analytical, confident, or cautious), they found that studies without commercial funding tended to use language that reflected more fear and a more impersonal tone than commercially funded studies. Langerhuizen et al. (2021) also analysed patients' online comments about healthcare providers and found that comments characterized by joy and confidence were associated with higher service ratings, whereas sadness and tentativeness were linked to lower evaluations.

This tool has also been applied in the field of music. Marouf et al. (2019), for instance, analysed a large corpus of English song lyrics and classified them according to both language style (analytical, confident, tentative) and emotional tone (anger, fear, joy, sadness). Extending this line of work, Somse et al. (2022) combined tone detection with voice analysis to identify users' emotional states and subsequently recommend music that matched their mood. More practical research has been done on neuromotor disability to help dependent people (Jain & Verma, 2020). The authors presented a solution to control the movement of people who speak clearly but cannot walk because of their disability, and proposed a machine-learning based methodology to detect emotion from speech to help people to interact better with their surroundings. Likewise, Gain and Hotti (2017) suggested that emotional tones and linguistic patterns extracted from text may also contribute to assessing personality traits and social tendencies. Taken together, these findings demonstrate the usefulness of Watson-based language analysis tools and indicate that they are sufficiently robust to capture emotional and linguistic variation in written discourse in domains as diverse as healthcare, social media, music, and scientific communication.

However, despite this potential, most previous applications of Watson tools have been developed outside the fields of philology and language teaching, which highlights the novelty of applying this tool to the study of emotional content in L2 written production. Nonetheless, a recent study (Sánchez et al., under review) applied this software to examine the effects of two instructional strategies—summary and guessing—on emotional writing performance in L2 English. In that study, learners' written productions before and after the intervention were analysed using Watson Tone

Analyzer to quantify the emotional tone of each text. Although the intervention did not produce statistically significant changes in overall emotional writing performance—measured as the average score of four emotional dimensions (anger, fear, joy, and sadness)—it did lead to a reduction in the analytical tone of the texts produced by both experimental groups. In this way, the tool made it possible to identify that both teaching strategies were similarly effective in encouraging learners to write in a less analytical and comparatively more affective manner.

Based on this rationale, the main aim of the present study was to further investigate possible changes in the emotional content of L2 written texts following specific instruction based on the summary strategy<sup>2</sup>. It was expected that the summary-based training would modulate the emotional content in L2 written production. Consequently, changes after training were expected to be reflected in both normative valence scores and sentiment scores. More specifically, the study aimed to test the validity of Watson NLU for measuring emotional language in written texts. It was hypothesized that sentiment scores (generated by Watson) would correlate with and predict mean normative valence scores (obtained from affective ratings traditionally used in L2 research) of the L2 written texts collected both before and after the teaching instruction.

## **2. Method**

### **2.1. Participants**

A group of Spanish undergraduate students (n=33, 7 males, Mage= 18.27) enrolled in the English Studies degree (University of Salamanca, Spain) participated in this study. They were B2-level English students (Council of Europe, 2018) who volunteered to participate in the research. This was done to inform participants about the study and to obtain permission to use their data anonymously and in aggregated form-administered to account for the study that was being carried out and to request permission so that their data could be used globally, never individually.

### **2.2. Procedure**

Participants wrote an autobiographical text (approximately 150 words) in their L2 (English) before (pretest) and after instruction (posttest) with emotional language (see section Instruction for more details).

#### **2.1.1. Pretest and Posttest**

In the pretest phase, students wrote about a dream (150 words) they had had in approximately 30 minutes. In the posttest phase, they wrote another short text (150 words) about a personal experience (30 minutes); this phase took place two weeks after the instruction sessions to measure long-term effects. The topics were chosen to generate texts in which participants felt inclined to use emotional vocabulary and expressions through the retelling of subjective autobiographical experiences (Pavlenko, 2012). These activities were administered online through the students' virtual campus.

### 2.1.2. Instruction

Instruction was provided in two 50-minute sessions one week apart, and the tasks participants completed during these two sessions were paper-based. In the first instruction session, a text adapted from a blog post was used to address negative, high-arousal emotions related to anger (see Appendix 1). Participants were first asked to read the text to become familiar with the words and emotional expressions, and then they were asked to summarize it. They were advised to outline the main ideas before summarizing the text to help them paraphrase and rephrase ideas. While writing the summary they could look at the text, and guidance and support were always provided by the instructor, in order to motivate the participants and make them feel more confident (Méndez López, 2016).

In the second instruction session, the input text used was adapted from a blog post on positive (pleasant) feelings (see Appendix 2). The procedure was the same as in the first session. Participants read and summarized the text, and while writing the summary they were allowed to look at the text and were encouraged to ask any questions.

The fact that the first text dealt with negative emotions (session 1) and the second with positive emotions (session 2) did not jeopardize the validity of the research because in the pre- and posttest participants were not directed towards positivity or negativity and could express themselves freely with the emotional terms learned in the instruction sessions.

### 2.3. Data analyses

A quasi-experimental pretest / posttest design (Larson-Hall, 2010; Rogers & Révész, 2020) was used to examine the effect of the teaching summary strategy on students' emotional L2 writing performance (cf., Sánchez et al., 2026, where this design was applied to test the effect of teaching strategies on vocabulary learning in EFL).

Students' texts were analyzed using two different yet complementary indices: lexical valence scores derived from a normative database and sentiment scores obtained through automated sentiment analyses by means of Watson NLU software tool. For the lexical valence analysis, each text produced in the pre- and post-instruction phases was first corrected for spelling and tokenized. Once all the words from each text were extracted, the content words (nouns, adjectives, verbs, and adverbs) were selected and lemmatized. Thus, words were reduced to their base form (e.g., singular nouns and infinitive verb forms) in order to facilitate matching with the normative database. Then, valence scores for each lexical item were obtained from the set of English affective norms provided by Warriner et al. (2013), which provides ratings on emotional valence of a large set of words, in on a scale ranging from 1 (very negative) to 9 (very positive). Nonetheless, content words that were not present in the database were excluded from the analysis (23.69% of the words in the pretest texts and 23.87% in the posttest texts). For each text in both phases, the mean valence score was computed by averaging the valence ratings extracted across all content words, providing an index of the affective polarity of the lexical items used in the text. This measure captures the emotional characteristics of the vocabulary, rather than the evaluative tone of the discourse as a whole. For descriptive purposes, the proportion of valenced words in each text was also extracted as an index of emotional vocabulary density. Following common practices in

previous studies, words with valence scores  $\geq 6$  and  $\leq 4$  were classified as emotional, and their proportion relative to the total number of words matched with the normative database was calculated for each text.

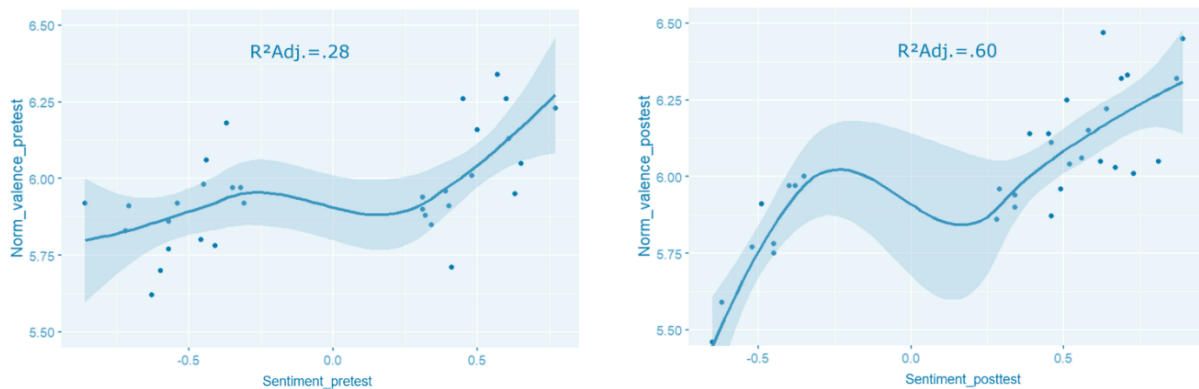
Regarding sentiment, scores for each text in pre- and posttest phases were automatically generated by Watson NLU tool. The system applies machine-learning models trained on large text corpora to analyze the linguistic features in the input text, providing a computational estimation for the emotional tone expressed in the text. The resulting sentiment scores range from -1 to 1, where values closer to 1 indicate a more positive tone, values closer to -1 reflect a negative tone, and values around 0 indicate a more neutral evaluative tone. Therefore, whereas lexical valence operates at the level of individual words, sentiment scores indicate the overall polarity of texts as a whole, taking into account the linguistic context in which words appear.

Then, two different analyses were carried out with both normative valence and sentiment scores obtained for the texts. First, to determine the relationship between sentiment and normative valence scores, correlational and regression analyses were carried out. Thus, Pearson correlations were computed between both sentiment and normative valence scores, separately for written texts obtained in pre- and posttest phases. Then, regression analyses were carried out considering sentiment scores as predictor or independent variable, and mean valence scores as dependent variable, separately conducted for texts written in the pre- and posttest phases. Second, to determine the effect of the specific training, paired-sample t-tests were performed to contrast written L2 texts in pre- and post-training phases, separately considering normative valence and sentiment scores. Statistical analyses were conducted with the SPSS package (IBM, version 23) and the R software (Core Team, 2021) was used to plot and visualize results by means of ggplot2 package (Wickham, 2016) implemented in R Studio (version 2022.02.0).

### 3. Results

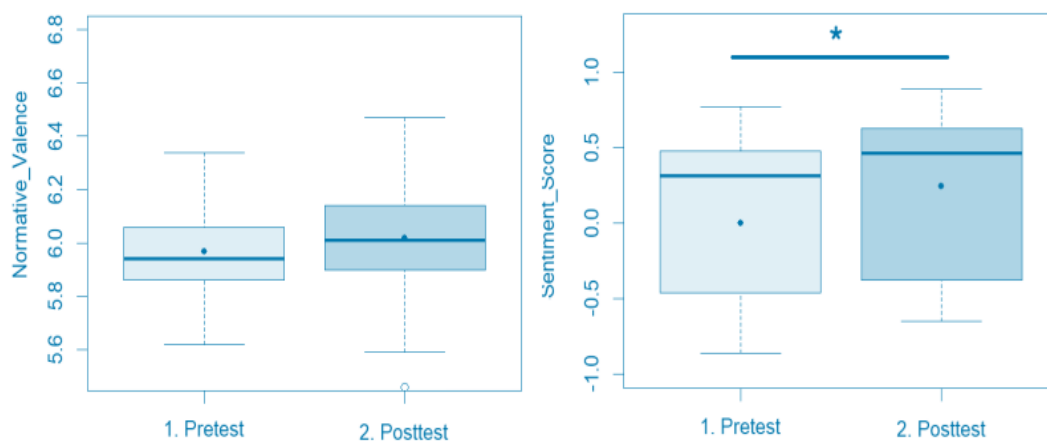
Correlational analyses conducted between sentiment and normative valence scores obtained in the pretest phase demonstrated a strong, positive relation between both indices ( $r=.55$ ,  $p=.001$ ). The relation was found even stronger when indices obtained in the posttest were analyzed ( $r=.78$ ,  $p<.001$ ).

Importantly, regression analyses confirmed that sentiment scores significantly predicted normative valence ratings, showing a strong, linear relation in the pretest [ $F(1, 32) = 13.9$ ,  $R^2\text{Adj.} = .28$ ,  $p = .001$ ; see Graph 1, left panel]. Moreover, such relation was found even stronger in the posttest phase [ $F(1, 32) = 50.21$ ,  $R^2\text{Adj.} = .60$ ,  $p < .001$ ; see Graph 1, right panel]. These results confirm that more positive sentiment scores actually predict more positive valence scores obtained through normative ratings, thus indicating Watson NLU as is able to determine the emotional tone of a given text.



Graph 1. Normative valence scores obtained for each written text in pretest (left panel) and posttest (right panel) as a function of sentiment scores (each point represents the mean obtained for each written text, for normative valence and sentiment scores)

Regarding the t-test carried out to compare the emotionality of L2 texts in pretest vs. posttest phases, no differences were observed in normative valence scores between both training phases [ $t(32) = 0.815$ ,  $p = 0.421$ , mean pretest = 5.966, mean posttest = 6.006, mean difference = -0.04]. See Graph 2 (left panel). Indeed, the proportion of emotionally valenced words was similar in both phases (pretest: 61.50%, posttest: 60.72%) and did not differ significantly across phases ( $p > .05$ ). Conversely, the analysis considering text-derived sentiment values revealed that the emotionality of the written texts significantly increased after the specific training [ $t(32) = -2.01$ ,  $p = .05$ , mean pretest = -0.002, mean posttest = 0.244, mean difference = -0.246]. See Graph 2 (right panel).



Graph 2. Distribution of normative valence scores (left panel) and sentiment scores (right panel) for L2 written texts across pretest and posttest phases. Dots within each boxplot represent the mean obtained in normative valence and sentiment scores for each pretest and posttest condition; the asterisk indicates significant differences for the contrast between sentiment scores in pretest and posttest phases

#### 4. Discussion

The first aim of the present study was to investigate whether specific training could promote changes in L2 emotionality in written production. Previous research has consistently reported differences between L1 and L2 in the processing and use of

emotional language, as shown through subjective evaluations and objective measures such as behavioral and neurophysiological responses (Caldwell-Harris, 2015; Ferré et al., 2010; Foroni, 2015; Kousta et al., 2009; Opitz & Degner, 2012). Since these differences are often attributed to L2 learning in affectively detached contexts, such as formal instructional settings, it was hypothesized that increasing learners' exposure to L2 words in emotional contexts through targeted training would encourage the use of emotional language and, consequently, improve emotional L2 communication.

To test this hypothesis, B2-level learners of English as an L2 underwent a training based on the summary strategy. Emotionality in their written production before and after the intervention was assessed using two complementary indices: normative valence scores derived from emotional norms in English (Warriner et al., 2013) and sentiment scores estimated through Watson NLU. Overall, results confirmed the usefulness of the training enhancing L2 emotionality in written production. Notably, this improvement was more clearly detected through discourse-level sentiment scores than through the traditional word-based approach based on normative valence ratings.

The emotional tone captured by sentiment scores significantly increased after the application of the summary strategy. This result supports the usefulness of this teaching approach for enhancing emotional expression in L2 written production and aligns with previous research showing the benefits of summarization for reading comprehension and vocabulary learning (Hsiang et al., 2020; Keck, 2014; Shokrpour et al., 2013; Stevens et al., 2019). Summarizing likely promotes deeper processing of affective language because learners must re-elaborate linguistic content through metacognitive operations such as identifying key information, paraphrasing, and synthesizing ideas during text production (Keck, 2006; Mokeddem & Houcine, 2016). This process may facilitate the integration of emotional L2 vocabulary into memory and improve later access to such vocabulary during writing. It is also possible that this strategy could strengthen learners' ability to express emotions in oral communication or even influence emotional experience in L2, although these possibilities remain open questions for future research.

However, contrary to our predictions, the emotional valence of the words used in the texts, measured through standardized normative ratings in English (Warriner et al., 2013), did not change significantly across testing phases, despite a slight increase after training. The different sensitivity shown by sentiment and valence scores may be explained by the distinct dimensions captured by each measure, particularly in a relatively small sample size. Whereas sentiment analysis evaluates the emotional tone of the text as a whole, emotional valence reflects the degree of pleasantness associated with individual lexical items, with the overall score calculated as the mean valence of the words produced. In this sense, sentiment analysis extends beyond the positive or negative connotations of isolated words and captures the emotional orientation emerging from discourse as a whole. This more context-sensitive and naturalistic perspective may therefore be especially appropriate for detecting changes in emotional expression in L2 writing. On a more interpretative level, this differential pattern might suggest that the pedagogical intervention influences how learners organized and conveyed emotional meaning at the text level rather than substantially modifying the emotional polarity of individual lexical items.

To our knowledge, this is the first study, together with Sánchez et al. (under review), to examine changes in written L2 emotionality following a specific teaching intervention. Previous research has mainly focused on differences in the number of emotional words produced in L2 texts, showing that learners with higher proficiency tend to use more emotionally marked vocabulary (Dewaele & Pavlenko, 2008; Kyriakou & Mavrou, 2023; Mavrou et al., 2025). Other studies have shown that teaching strategies involving emotional engagement can help learners build emotional associations with new vocabulary, which supports vocabulary retention (Pishghadam & Shayesteh, 2016). Positive emotional stimuli have also been used to facilitate L2 vocabulary learning (Kralova et al., 2022), although changes in the emotionality of new vocabulary gained through the intervention were not specifically tested. In addition, teaching approaches based on multimodal elaborative processing—such as imagination, visual and spoken language, body expression, and gestures—have been found to improve emotional vocabulary learning in the EFL classroom (Sánchez et al., 2026) and to influence learners' perception of emotional word connotations (Sánchez et al., 2025).

A second objective of the present study was to test the usefulness of Watson NLU as a tool for measuring emotional language in L2 written texts. To address this aim, the study moved beyond the traditional word-based approach commonly used in previous research (e.g., Kousta et al., 2009; Opitz & Degner, 2012; Palazova et al., 2011) by analyzing emotionality at the discourse level, that is, considering the text as a whole rather than considering the valence of isolated lexical items. This made it possible to assess whether sentiment scores generated by Watson NLU converged with mean normative valence scores derived from affective ratings traditionally used in L2 research. Results showed that sentiment scores not only correlated with, but also predicted, the mean emotional valence of the words used in the texts, particularly after the training. This finding supports the hypothesis that sentiment analysis is sensitive to variations in emotional language and provides internal validation for the sentiment-based approach adopted in this study. In this sense, the convergence between sentiment and normative valence indices supports the validity of Watson NLU as a reliable tool for capturing emotional language in L2 written discourse and for detecting changes associated with instructional intervention. Furthermore, these findings extend previous research conducted in other fields showing the utility of AI-based language analysis tools for characterizing emotional aspects of written language (Gain & Hotti, 2017; Jain & Verma, 2020; Langerhuizen et al., 2021; Maleki et al., 2023; Marouf et al., 2019; Somse et al., 2022; Steffens et al., 2021), and suggest that such tools can also capture variation in emotional expression associated with language learning experiences.

Taken together, these findings contribute both to pedagogical research on how instructional strategies can enhance emotional expression in L2 writing and to methodological research by supporting the use of AI-based sentiment analysis tools for studying emotional language in L2 contexts. From an educational perspective, the results suggest that tools capable of evaluating emotional tone in written discourse may help teachers carry out more objective assessments of students' written production beyond the sentence level. In addition, automated analysis considerably reduces the time and effort typically required to evaluate student texts. More broadly, this type of

naturalistic assessment may help researchers further investigate L1–L2 differences in emotional processing. Future studies may use this AI tool or similar ones to further explore their potential and confirm the results obtained here. Additionally, future research could examine whether sentiment analysis tools interpret emotional tone differently across genders, which may help identify potential biases in automated assessments and contribute to a more robust and equitable use of these technologies.

### Acknowledgments

This research was supported by Grant PID2021-122465NB-I00 funded by MCIN/AEI/10.13039/501100011033 and by “ERDF A way of making Europe.”

### Notes

1. Watson Natural Language Understanding is the equivalent to Watson Tone Analyzer, which is no longer available after 24 February 2023 (<https://www.ibm.com/demos/live/natural-language-understanding/self-service>).
2. Part of the texts produced by students, selected from a larger dataset belonging to a study currently under review (Sánchez et al., under review), were compiled and analyzed between March and May 2023.

### References

- Caldwell-Harris, Catherine L. (2015). Emotionality differences between a native and foreign language. *Current Directions in Psychological Sciences*, 24(3), 214-219. <https://doi.org/10.3389/fpsyg.2014.01055>
- Conrad, Markus, Recio, Guillermo, & Jacobs, Arthur M. (2011). The time course of emotion effects in first and second language processing: A cross cultural ERP study with German-Spanish bilinguals. *Frontiers in Psychology*, 2, 351. <https://doi.org/10.3389/fpsyg.2011.00351>
- Council of Europe. (2018). *Common European framework of reference for languages: Learning, teaching, assessment. Companion volume with new descriptors*. Cambridge University Press. [www.coe.int/lang-cefr](http://www.coe.int/lang-cefr)
- Dewaele, Jean-Marc. (2004). The emotional force of swearwords and taboo words in the speech of multilinguals. *Journal of Multicultural and Multilingual Development*, 25, 204–222. <https://doi.org/10.1080/01434630408666529>
- Dewaele, Jean-Marc, & Pavlenko, Aneta. (2008) Emotion vocabulary in interlanguage. *Language Learning*, 52(2-3), 263–322. <https://doi.org/10.1111/0023-8333.00185>
- Eilola, Tiina M., & Havelka, Jelena. (2010). Affective norms for 210 British English and Finnish nouns. *Behavior Research Methods*, 42(1), 134–140. <https://doi.org/10.3758/BRM.42.1.134>

- Emir Özder, Levent, Özer, Demet, & Göksun, Tilbe. (2023). Gesture use in L1-Turkish and L2-English: Evidence from emotional narrative retellings. *Quarterly Journal of Experimental Psychology*, 76(8), 1797-1816. <https://doi.org/10.1177/17470218221126685>
- Fan, Angela, Lewis, Mike, & Dauphin, Yann. (2018). Hierarchical neural story generation. *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics* (pp. 889–898). Association for Computational Linguistics.
- Ferré, Pilar, García, Teófilo, Fraga, Isabel, Sánchez-Casas, Rosa, & Molero, Margarita. (2010). Memory for emotional words in bilinguals: Do words have the same emotional intensity in the first and in the second language? *Cognition and Emotion*, 24(5), 760–785. <https://doi.org/10.1080/02699930902985779>
- Froni, Francesco. (2015). Do we embody second language? Evidence for ‘partial’ simulation during processing of a second language. *Brain and Cognition*, 99, 8-16. <https://doi.org/10.1016/j.bandc.2015.06.006>
- Gain, Ulla, & Hotti, Virpi. (2017). Tones and traits - experiments of text-based extractions with cognitive services. *Finnish Journal of EHealth and EWelfare*, 9(2-3), 82–94. <https://doi.org/10.23996/fjhw.61001>
- Harris, Catherine L. (2004). Bilingual speakers in the lab: Psychophysiological measures of emotional reactivity. *Journal of Multicultural and Multilingual Development*, 25(2-3), 223–247. <https://doi.org/10.1080/01434630408666530>
- Harris, Catherine L., Gleason, Jean Berko, & Ayçiçeği, Ayşe. (2006). When is a first language more emotional? Psychophysiological evidence from bilingual speakers. In Aneta Pavlenko (Ed.), *Bilingual minds: Emotional experience, expression, and representation* (pp. 257–283). Multilingual Matters.
- Hsiang, Tien Ping, Graham, Steve, & Yang, Yu-Mao. (2020). Teachers’ practices and beliefs about teaching writing: A comprehensive survey of grades 1 to 3 teachers. *Reading and Writing*, 33, 2511-2548. <https://doi.org/10.1007/s11145-020-10050-4>
- IBM Watson Natural Language Understanding (demo). (<https://www.ibm.com/demos/live/natural-language-understanding/self-service>)
- IBM Watson Natural Language Understanding (information). [https://cloud.ibm.com/registration?target=/catalog/services/natural-language-understanding%3FhideTours%3Dtrue%26&cm\\_sp=WatsonPlatform-WatsonPlatform--OnPageNavCTA-IBMWatsonNaturalLanguageUnderstanding--WatsonDeveloperWebsite](https://cloud.ibm.com/registration?target=/catalog/services/natural-language-understanding%3FhideTours%3Dtrue%26&cm_sp=WatsonPlatform-WatsonPlatform--OnPageNavCTA-IBMWatsonNaturalLanguageUnderstanding--WatsonDeveloperWebsite)
- Imbault, Constance, Titone, Debra, Warriner, Amy Beth, & Kuperman, Victor. (2021). How are words felt in a second language: Norms for 2,628 English words for valence and arousal by L2 speakers. *Bilingualism: Language and Cognition*, 24(2), 281-292. <http://doi.org/10.1017/S1366728920000474>
- Jain, Manish, & Verma, Madhushi. (2020). Automated motion of a robot based on emotion analysis. *Journal of Physics: Conference Series*, 1438, 012013. <https://doi.org/10.1088/1742-6596/1438/1/012013>

- Kazanas, Stephanie. A., & Altarriba, Jeanette. (2016). Emotion word type and affective valence priming at a long stimulus onset asynchrony. *Language and Speech*, 59(3), 339-52. <http://doi.org/10.1177/0023830915590677>
- Keck, Casey. (2006). The use of paraphrase in summary writing: A comparison of L1 and L2 writers. *Journal of Second Language Writing*, 15(4), 261-278. <https://doi.org/10.1016/j.jslw.2006.09.006>
- Keck, Casey. (2014). Copying, paraphrasing, and academic writing development: A re-examination of L1 and L2 summarization practices. *Journal of Second Language Writing*, 25, 4-22. <https://doi.org/10.1016/j.jslw.2014.05.005>
- Khoshsima, Hooshang, & Rabani Nia, Maryam. (2014). Summary strategies and writing ability of Iranian intermediate EFL students. *International Journal of Language and Linguistics*, 2(4), 263-272. <https://doi.org/10.11648/j.ijll.20140204.14>
- Kousta, Stravoula Thaleia, Vinson, David P., & Vigliocco, Gabriella. (2009). Emotion words, regardless of polarity, have a processing advantage over neutral words. *Cognition*, 112(3), 473-81. <http://doi.org/10.1016/j.cognition.2009.06.007>
- Kralova, Zdena, Kamenicka, Jana, & Tirpakova, Anna. (2022). Positive emotional stimuli in teaching foreign language vocabulary. *System*, 104(1), 102678. <https://doi.org/10.1016/j.system.2021.102678>
- Kyriakou, Andreas, & Mavrou, Irini. (2023). ¿Eres muy emocional? I don't think so. How does language influence our emotional responses to everyday moral dilemmas? In A. Blanco Canales & S. Martín Leralta (Eds.). *Emotion and identity in foreign language learning* (pp. 297-321). Peter Lang.
- Kyriakou, Andreas, Mavrou, Irini, & Palapanidi, Kiriakí. (2024). The role of foreign language in the experience and emotional expression of guilt: Evidence from moral scenarios and autobiographical memories of bilinguals. *International Journal of Bilingual Education and Bilingualism*, 27(10), 1407-1421. <https://doi.org/10.1080/13670050.2024.2365238>
- Lam, Boji P. W., & Mardquart, Thomas P. (2022). Emotional and non-emotional verbal fluency in native and non-native speakers. *Archives of Clinical Neuropsychology*, 37(1), 199-209. <https://doi.org/10.1093/arclin/acab031>
- Langerhuizen, David W. G., Brown, Laura E., Doornberg, Job N., Ring, David, Kerkhoffs, Gino M. M. J., & Janssen, Stein J. (2021). Analysis of online reviews of orthopaedic surgeons and orthopaedic practices using natural language processing. *Journal of the American Academy of Orthopaedic Surgeons*, 29(8), 337-344. <https://doi.org/10.5435/JAAOS-D-20-00288>
- Larson-Hall, Jenifer. (2010). *A guide to doing statistics in second language research using SPSS*. Routledge.
- Maleki, Negar, Padmanabhan, Balaji, & Dutta, Kaushik. (2023). The effect of monetary incentives on health care social media content: Study based on topic modeling and sentiment analysis. *Journal of Medical Internet Research*, 25, e44307. <https://doi.org/10.2196/44307>
- Marouf, Ahmed Ali, Hossain, Rafayet, Kabir Rasel Sarker, Md Rahmatul, Pandey, Bishwajeet, & Tanvir Siddiquee, Shah. Md. (2019). Recognizing language and emotional tone from music

- lyrics using IBM Watson Tone Analyzer. In *2019 IEEE International Conference on electrical, computer and communication technologies (ICECCT)*.  
<https://doi.org/10.1109/ICECCT.2019.8869008>
- Mavrou, Irini, Bustos, Fernando, & Chao, Javier. (2025). Emotional vocabulary in immigrants' L2 written discourse: Is linguistic distance a proxy for L2 emotionality? *Journal of Multilingual and Multicultural Development*, 46(8), 2325-2341.  
<https://doi.org/10.1080/01434632.2023.2284894>
- Méndez López, Mariza. (2016). Las emociones en el aprendizaje de una lengua extranjera: su impacto en la motivación. *Revista Internacional de Lenguas Extranjeras*, 5, 27-46.  
<https://doi.org/10.17345/rile5.1002>
- Mokeddem, Samiha, & Houcine, Samira. (2016). Exploring the relationship between summary writing ability and reading comprehension: Toward an EFL writing-to-read instruction. *Mediterranean Journal of Social Sciences*, 7(2), 197-205.  
<https://doi.org/10.5901/mjss.2016.v7n2s1p197>
- Opitz, Bertram, & Degner, Juliane. (2012). Emotionality in a second language: It's a matter of time. *Neuropsychologia*, 50(8), 1961-1967.  
<https://doi.org/10.1016/j.neuropsychologia.2012.04.021>
- Palazova, Marina, Matwill, Katharina, Sommer, Werner, & Schacht, Annkathrin. (2011). Are effects on emotion in single words non lexical? *Neuropsychologia*, 49(9), 2766-75.  
<http://doi.org/10.1016/j.neuropsychologia.2011.06.005>
- Pavlenko, Aneta. (2005). *Emotions and bilingualism*. Cambridge University Press.
- Pavlenko, Aneta. (2012). Affective processing in bilingual speakers: Disembodied cognition? *International Journal of Psychology*, 47(6), 405-428.  
<https://doi.org/10.1080/00207594.2012.743665>
- Pavlenko, Aneta, & Driagina, Viktoria. (2007). Russian emotion vocabulary in American learners' narratives. *Modern Language Journal*, 91(2), 213-234. <https://www.jstor.org/stable/4626001>
- Pérez-García, Elisa., & Sánchez, María Jesús. (2020). Emotions as a linguistic category: Perception and expression of emotions by Spanish EFL students. *Language Culture and Curriculum*, 33(3), 274-289. <https://doi.org/10.1080/07908318.2019.1630422>
- Pishghadam, Reza, & Shayesteh, Shaghayegh. (2016). Emotioncy: A post-linguistic approach toward vocabulary learning and retention. *Sri Lanka Journal of Social Sciences*, 39(1), 27-36. <http://dx.doi.org/10.4038/sljss.v39i1.7400>
- R Core Team (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
- Rogers, John, & Révész, Andrea. (2020). Experimental and quasi-experimental designs. In Jim McKinley & Heath Rose (Eds.), *The Routledge handbook of research methods in applied linguistics* (pp. 133-143). Routledge.
- RStudio Team (2020). *RStudio: Integrated Development for R*. RStudio, PBC, Boston, MA.  
<http://www.rstudio.com/>

- Sánchez, María Jesús, Fernández, María, & Pérez-García, Elisa. (2026). Comparison of the context provision and imagination elicitation approaches to learning emotional vocabulary. *Language Teaching Research*, 30(2), 693-715. <https://doi.org/10.1177/13621688221131449>
- Sánchez, María Jesús, Pérez-García, Elisa, López, Belén, & Bermúdez-Margaretto, Beatriz. (2025, published online). Bridging the gap between L1 and L2: Enhanced emotional vocabulary through elaborative processing in Spanish-speaking English Language learners. *International Journal of Applied Linguistics*. <https://doi.org/10.1111/ijal.70064>
- Sánchez, María Jesús, Pérez-García, Elisa, & Bermúdez-Margaretto, Beatriz. (under review). Production of written texts with emotional features in EFL: Summary or guessing strategies?
- Sheikh, Naveed A., & Titone, Debra. (2016). The embodiment of emotional words in a second language: An eye-movement study. *Cognition and Emotion*, 30(3), 488-500. <http://doi.org/10.1080/02699931.2015.1018144>
- Shokrpour, Nasrin, Sadeghi, Azin, & Seddigh, Fatemeh. (2013). The effect of summary writing as a critical reading strategy on reading comprehension of Iranian EFL learners. *Journal of Studies in Education*, 3(2), 127-138. <https://doi.org/10.5296/jse.v3i2.2644>
- Somse, Shubham, Kulkarni, Amruta, & Chaugule, Balaji. (2022). Music room - A chatbot based song recommender using sentiment analysis. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, 8(3), 514-517. <https://ijsrcseit.com/home/issue/view/article.php?id=CSEIT228506>
- Steffens, Anath N. V., Langerhuizen, David W. G., Doornberg, Job N., Ring, David, & Janssen, Stein J. (2021). Emotional tones in scientific writing: Comparison of commercially funded studies and non-commercially funded. *Acta Orthopaedica*, 92(2), 240-243. <https://doi.org/10.1080/17453674.2020.1853341>
- Stevens, Elizabeth A., Park, Sunyoung, & Vaughn, Sharon. (2019). A review of summarizing and main idea interventions for struggling readers in grades 3 through 12: 1978-2016. *Remedial and Special Education*, 40(3), 131-149. <https://doi.org/10.1177/0741932517749940>
- Taherdoost, Hamed, & Madanchian, Mitra. (2023). Artificial intelligence and sentiment analysis: A review in competitive research. *Computers*, 12(2), 37. <https://doi.org/10.3390/computers12020037>
- Tang, Enze, & Ding, Hongwei. (2024). Emotion effects in second language processing: Evidence from eye movements in natural sentence reading. *Bilingualism: Language and Cognition*, 27(3), 460-479. <http://doi.org/10.1017/S1366728923000718>
- Toivo, Wilhelmiina, & Scheepers, Christoph. (2019). Pupillary responses to affective words in bilinguals' first versus second language. *PLoS ONE*, 14(4), e0210450. <https://doi.org/10.1371/journal.pone.0210450>
- Vidal Noguera, Carmen, & Mavrou, Irini. (2025). The 'emotional brain' of adolescent Spanish-German heritage speakers: is emotional intelligence a proxy for productive emotional vocabulary? *Bilingualism: Language and Cognition*, 28(1), 217-231. <https://doi.org/10.1017/S1366728924000348>

Warriner, Amy Beth, Kuperman, Victor, & Brysbaert, Marc. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods*, 45(4), 1191–1207. <http://doi.org/10.3758/s13428-012-0314-x>

Wickham, Hadley. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. <https://ggplot2.tidyverse.org>

Winkel, Heather. (2013). The emotional Stroop task and emotionality rating of negative and neutral words in late Thai–English bilinguals. *International Journal of Psychology*, 48(6), 1090-1098. <http://doi.org/10.1080/00207594.2013.793800>

## Contributions according to the Credit system

Conceptualization: MJS; Literature review: MJS, EPG; Methodology: EPG, BBM; Validation: MJS, BBM; Data management: MJS, BBM; Drafting of the original manuscript: MJS, EPG, BBM; Review and editing: MJS, EPG.

## Anexo I.

Text used in the 1st instruction session (full version)

TEXT

(Adapted from the blog article “[Things to do when you’re feeling angry with someone](#)”)

We all have lots of misunderstandings and annoyances (e.g., I felt angry because I have always struggled with saying n. felt misunderstood and judged). With this in mind, I put together this guide to dealing with anger.

SIT WITH YOUR ANGER

1. Allow yourself to feel angry. You may think you need to cover “negative feelings” with positive ones. You don’t. You’re entitled to feel whatever you need to feel. We all are.
2. Feel the anger in your body. Is your neck tense? Is your chest burning? Is your throat tightening? Are your legs twitching? Recognize the sensations in your body and breathe into those areas to clear the blockages that are keeping you feeling stuck.

EXPLORE YOUR ANGER

3. Check in with your mood before the incident. Were you having a bad day already? Were you already feeling annoyed or irritated? It could be that someone’s actions were the straw that broke the camel’s back, but not fully responsible for creating these feelings.
4. Ask yourself: Why is this bothering you so much? Is it really what someone else did, or are you feeling angry because of what you’re interpreting their actions to mean? (For example, you may think that your boyfriend not showing up means that he doesn’t respect you, when he may have a valid explanation).
5. Put it in a letter. Now that you know more clearly what part the other person played in your anger and which part is more about you, write a letter to him or her. You may send this letter, or you might end up just burning it. This is to help you clarify what exactly you’d like that person to know, understand, or change.

### RESPOND WITHOUT ANGER

6. Now that you're clear about the role you played in your anger, initiate a verbal conversation about what bothered you. You could also send the letter you wrote, but it will be easier to clarify parts the other person doesn't understand if you're having a direct back-and-forth exchange.

7. Use "I feel" language. So instead of saying, "You didn't show up so you obviously don't care about me," say, "When you forget about the things that are important to me, I feel hurt." In this way, you're not assuming the other person meant to make you feel bad—you're just explaining how it makes you feel so they can understand how their actions impact you.

8. Focus on creating a solution. If your goal is to get the other person to admit that they're wrong, you'll probably end up in a power struggle. Focus instead on what you'd like to change in the future—for example, you'd appreciate it if she would come straight to you next time instead of complaining about you behind your back. You can help facilitate this by owning some responsibility—that you will listen if he comes to you instead of getting emotional.

### LEARN FROM YOUR ANGER

9. Learn what you value. This situation taught you something useful about what you value in the people you choose to be friends with—maybe directness, humility, or loyalty. This will help you decide which people you might want to spend more or less time with going forward.

10. Learn how to communicate clearly. This experience was an exercise in expressing yourself in the best way to be heard and understood. There will definitely be more situations like this in the future, so this is good practice for misunderstandings and struggles to come.

11. Learn how you can improve your response to anger going forward. Maybe you reacted too quickly, so now you've learned to put more space between your feelings and your response. Maybe you got defensive, and the other person shut down, so you've learned to be less accusatory in the future.

## Anexo II.

Text used in the 2nd instruction session (full version)

### TEXT

(Adapted from the blog article "[What are the ten positive actions?](#)")

One of my favorite books to come out of the "positive psychology" movement is called *Positivity*, by Dr. Barbara Fredrickson. Truly a genius and pioneer in the field, Dr. Fredrickson has been studying positive emotions in her lab long before it was vogue. Her data reveals that negative emotions, like fear, can close down our ability to function, while positive emotions open us up to possibility, and an increased ability to move forward.

She prefers the term "Positivity" to "Happiness" and stresses the importance and possibility of not just being happy; but flourishing. Isn't that a lovely word? Wouldn't we all love to flourish? Check out Dr. Fredrickson herself describing "Positivity" and why it is so important at this moment in history. As she says at the end of the clip: "Investing in things that bring us more positive emotions is an investment in our future. Choosing Hope over Fear." Dr. Fredrickson's came up with a top 10 list of positive emotions, in order of most frequent to least. Allow yourself an opportunity to scroll through the list and ask yourself, "When did I last fully experience this emotion?" The answers may surprise you.

1. Joy happens in an instant -- a perfect moment captured when all is just exactly as it should be. Think of a wonderful holiday morning with the family, an unexpected present that delights you, or seeing the first smile on your infant's face. What brings you Joy?

2. Gratitude is a moment of realizing someone has gone out of their way for you, or simply feeling overwhelmed with your heart opening, after being moved in some way. With gratitude comes a desire to give in return or 'pay it forward' in some way. When did you last experienced deep Gratitude?

3. Serenity is like a mellow, relaxed, or sustained version of Joy. Serenity is a peacefulness that comes on a cloudless day, when you realize there's nothing you have to do. Serenity is indulging in a favorite luxury, and being mindful enough to take it in. Serenity is the moment on vacation when you finally let go. Has Serenity crossed your door lately?

4. Interest is a heightened state that calls your attention to something new that inspires fascination, and curiosity. Like a shiny new toy to capture your imagination, interest is alive and invigorating. Interest wakes you up, and leaves you wanting more. What Interests you these days?

5. Hope. Dr. Fredrickson describes it best: "Unlike other emotions that arise out of comfort and safety, hope springs out of dire circumstances, as a beacon of light. Deep within the core of hope is the belief that things can change, turn out better. Possibilities exist. Hope sustains you and motivates you to turn things around." The inauguration of President Obama brought me Hope. What brings you Hope?

6. Pride. Ever done something really well that took a little time and effort? Maybe you reached a goal you never thought was attainable? Then pat yourself on the back with unadulterated Pride. Stand back, take that deep breath and let it in -- you earned it. What have you done that made your proud?

7. Amusement. Think of amusement as those delightful surprises that make you laugh. It's those unexpected moments that interrupt your focus and crack you up. It's a great feeling to have amusement sparkle out of the doldrums and instantly change your perspective. Have you had any amusement in your life recently?

8. Inspiration is a moment that touches your heart and nearly takes your breath away -- or takes in your breath, as the word literally translates. Inspiration whispers between the strands of your hair, as you watch a perfect sunset, witness academic or athletic excellence, or observe unexpected triumphs over adversity. What brings Inspiration in your life?

9. Awe happens when you come across goodness on a grand scale, and you feel overwhelmed by greatness. Awe is triggered when we are faced with the vastness of Nature, or the cosmos. Gazing at the Milky Way and counting the stars or standing at the top of the Grand Canyon triggers awe. Have you had a moment of awe lately?

10. Love. The #1 most frequent positive emotion is here at the bottom. Love encompasses all of the above: joy, gratitude, serenity, interest, hope, pride, amusement, inspiration and even awe. When we experience love, our bodies are flooded with the "feel good" hormones that reduce stress and even lengthen our lives.