



UNIVERSIDAD DE COSTA RICA



Jorge Antonio Leoni de León y Édgar Casasola Murillo,
investigadores del Instituto de Investigaciones
Lingüísticas

Foto: [Laura Rodríguez Rodríguez](#)

IX Jornadas de Investigación del Instituto de Investigaciones Lingüísticas

Soltemos la lengua: Emociones, lenguaje y contexto en inteligencia artificial

Imaginemos un sistema en el que sería posible hacer búsquedas por emociones, más que por términos específicos

20 DIC 2024 Artes y Letras

¿Cuál es la relación entre las palabras y las emociones? ¿Cómo podemos extraer información con carga emocional a partir de un corpus? ¿Qué parámetros podemos emplear para determinar la relación entre las emociones y lo dicho? Incluso, ¿podemos establecer alguna especie de estructura de las emociones? ¿Sería esta una estructura específica a cada lengua? ¿Sería posible utilizar esa estructura en sistemas de Procesamiento del Lenguaje Natural (PLN)? A fin de responder a estas y otras preguntas relevantes para la Inteligencia Artificial (IA), nos hemos dado a la tarea de construir una taxonomía de términos relativos a emociones para el español. Propuestas similares existen muchas para el inglés. Un método en boga es utilizar una estructura ya existente en esa lengua y traducirla directamente al español. Sin embargo, los términos del inglés relativos a emociones no se correlacionan uno a uno con los del español, porque no es una imagen especular: la manera en que cada lengua construye sus categorías para referirse a la realidad puede ser muy diferente y varía según diversos factores, como el contexto, el tiempo y el lugar. Por ejemplo, el término *gladness* puede ser traducido en español como *regocijo*, *agradecimiento* o *satisfacción*. Del español al inglés, también podemos mencionar la palabra *fascinación* que tiene varias correspondencias, como *thrill* o

enthrallment. Esto según datos consultados con hablantes bilingües. También la información extraída por medio de *Grandes Modelos del Lenguaje* (LLMs por sus siglas en inglés) apunta en ese sentido. Es importante apuntar aquí que en nuestra investigación utilizamos ChatGPT 4.0, Gemini y Claude. Para el término *glumness*, ChatGPT indicó la traducción *melancolía*; Gemini produjo *tristeza*; y Claude propuso *abatimiento*. Ninguna de estas traducciones es errónea, sino que dependen del contexto y de la estructura semántica en acción, que en nuestro caso, es la del español. Por esta razón llegamos a la conclusión de que debíamos construir una taxonomía de términos relacionados con las emociones que responda a las características de la lengua española para este momento y este lugar.

Así, en el caso de Costa Rica, es necesario tomar en cuenta que una persona de este país puede decir “me siento pura vida”, frase que expresa emoción y que no es inmediatamente comprensible en todo el mundo hispanohablante. Aunque un humano podría detectar la emoción en esa frase, una aplicación de IA no necesariamente lo logra, además de que en algunos casos ciertas expresiones podrían tener connotaciones diferentes e incluso opuestas.

Una taxonomía como la que proponemos es muy útil en IA, por cuanto permite asociar frases con emociones. Imaginemos un sistema en el que sería posible hacer búsquedas por emociones, más que por términos específicos. También podríamos pensar en aplicaciones que identifiquen el grado de emoción en un discurso o en un artículo periodístico; más aún, debería ser posible establecer el balance de emociones presentes en un documento, algo que sería sumamente valioso para el análisis de opiniones, área de investigación muy activa con aplicaciones prácticas en mercadeo, política, deportes y entretenimiento.

Para construir un modelo que pueda evaluar la carga emocional de un texto, es necesario identificar las palabras asociadas con emociones *denominadas* primarias. Las emociones de primer nivel en nuestra taxonomía son *amor*, *alegría*, *ira*, *tristeza* y *miedo*. Es interesante que en inglés se incluye una sexta, que es la *sorpresa*, sin embargo en nuestra investigación no fue necesario incluirla. Nuestra taxonomía fue construida con base en dos fuentes. La primera fuente corresponde a los trabajos de Shaver (1987) y Parrot (2010), que proponen una estructura jerárquica de tres niveles. La segunda fuente es documental y consiste en la extracción de términos asociados a emociones a partir de un corpus representativo del español de Costa Rica. Rápidamente nos dimos cuenta, como ya lo mencionamos, que la estructura semántica del español referida a emociones es distinta al inglés, de manera que debíamos proponer una jerarquía específica. Así, en la taxonomía que proponemos, de manera similar al inglés, las emociones de 2o nivel se derivan de las de primer nivel y las de 3er nivel, de las de 2o nivel. Sin embargo, los términos referidos a emociones en español se agrupan de manera ligeramente diferente, de manera que la estructura léxica resultante es específica al español. Además, postulamos un 4o nivel que resulta de la interacción entre las emociones de los otros niveles, con diferentes grados de intensidad. Por ejemplo, el término *compasión* está asociado con *amor* (1er nivel) y con *tristeza* (también de 1er nivel). Más que una taxonomía, nuestra clasificación es una red de palabras interrelacionadas. Esta estructura es crucial para aplicaciones de IA, ya que ayuda a reducir la ambigüedad en la interpretación de emociones y mejora la precisión en la identificación de los matices emocionales. Además, otro aporte importante es que la construcción de la *taxonomía* está asociada con frases y frecuencias de uso léxico, por lo que tenemos las apariciones de cada término y los contextos en que se producen. Esta es una investigación basada en datos, que proporciona evidencia científica del uso de las palabras asociadas a emoción y a partir de las cuales inferimos una estructura. Uno de los resultados consiste en estadísticas que permiten determinar la prototipicidad, que podemos definir en nuestra investigación como la representatividad de un término con respecto a la emoción que evoca. Para la construcción de la taxonomía utilizamos un conjunto de datos de texto escrito en español (http://www.earthlings.io/download_cglu.html) que consta de 284 megabytes de texto

mayormente escrito en español costarricense (no es posible garantizar al 100 % el origen geográfico de los documentos).

Para la IA, esta adaptación lingüística ayuda a mejorar el reconocimiento emocional en aplicaciones que procesan texto en español y que incluyen desde chatbots hasta sistemas de análisis de sentimiento. A fin de conseguirlo empleamos *word embeddings*, que es una técnica que convierte palabras en vectores matemáticos para analizar la proximidad de términos y su evolución en el tiempo. Esto es especialmente útil para reflejar cómo cambian las asociaciones emocionales con ciertos términos según el contexto. En cuanto a las aplicaciones, un modelo robusto y confiable, como el que proponemos, que se apega a las especificidades de la lengua española, repercute en la calidad de las herramientas de software que eventualmente se implementen.

Además de mejorar el análisis y procesamiento de texto en español, una taxonomía como la nuestra aporta datos más precisos sobre el contenido emocional expresado en diferentes tipos de documentos (textos cortos, artículos periodísticos, opiniones en redes sociales o publicidad). Por ejemplo, una empresa que ha lanzado un producto nuevo puede verse en la situación de decidir si debe o no retirarlo del mercado; para este efecto sería valioso contar con información sobre el contenido emocional de las opiniones emitidas por los clientes. De ahí que nuestra investigación proporciona un valioso recurso para el desarrollo de aplicaciones científicas, gubernamentales y comerciales, relacionadas con el mejoramiento de servicios y estimación del grado de satisfacción de las personas.

Referencias

Shaver, P., Schwartz, J., Kirson, D., & O'connor, C. (1987). Emotion knowledge: further exploration of a prototype approach. *Journal of personality and social psychology*, 52(6), 1061.

Gerrod Parrott, W. (2010). Ur-emotions and your emotions: Reconceptualizing basic emotion. *Emotion review*, 2(1), 14-21.

“Soltemos la lengua” es una sección del proyecto *Esta palabra es mía*, un espacio de divulgación lingüística y literaria.

[Édgar Casasola Murillo](#)

Investigador del Instituto de Investigaciones Lingüísticas

edgar.casasola@ucr.ac.cr

[Jorge Antonio Leoni de León](#)

Investigador del Instituto de Investigaciones Lingüísticas

antonio.leoni@ucr.ac.cr

Etiquetas: [esta palabra es mia](#), [soltemos la lengua](#), [inteligencia artificial](#), [ia](#), [emociones](#), [lenguaje](#).

