

Análisis emocional del corpus COLUMNAS.HUMOR: un enfoque mixto

Chelo Vargas-Sierra¹

Recibido: 27 de enero de 2022 / Aceptado: 13 de junio de 2023

Resumen. El estudio de las emociones y la polaridad en el lenguaje ha adquirido importancia en los últimos años debido a su relevancia en áreas como la inteligencia artificial o el análisis de sentimientos en redes sociales, publicidad y comunicación en general, pues busca entender el modo en que las personas interactúan y se relacionan a través del lenguaje. Los estudios de este tipo son fundamentales para poder interpretar los sentimientos y las actitudes de los usuarios y clientes, así como para mejorar la interacción humano-máquina y la experiencia del usuario en distintos ámbitos. La polaridad y las emociones en el lenguaje se hace especialmente relevante en el análisis del humor, ya que la comprensión de sus patrones emocionales puede ser útil en el desarrollo de sistemas de inteligencia artificial capaces de detectar, entender y generar humor de manera efectiva. El presente estudio analiza los datos del corpus COLUMNAS.HUMOR (textos completos y secuencias humorísticas) con el objetivo de observar la polaridad y las emociones predominantes. La metodología utilizada está basada en corpus, con aplicación de técnicas cuantitativas y cualitativas y con el uso de herramientas informáticas para extraer datos de polaridad y de emociones. A partir de los diferentes tipos de resultados frecuenciales obtenidos, se realiza un análisis cualitativo para interpretar los datos en cada corpus de estudio. Este exhaustivo análisis nos ha permitido detectar las diferencias en el sentimiento y la verbalización emocional, así como contrastar los resultados por corpus y por el género de los periodistas estudiados.

Palabras clave: análisis de sentimiento, emociones, humor, género.

[en] Emotional analysis of the COLUMNAS.HUMOR corpus: A mixed approach

Abstract. The study of emotions and polarity in language has been the focus of increased attention over recent years due to its relevance in areas such as artificial intelligence and the analysis of sentiment on social networks, advertising, and communication in general, given that it seeks to understand how people interact and relate through language. Studies of this type are fundamental towards being able to interpret the feelings and attitudes of users and customers, as well as to improve human-machine interaction and the user experience in a variety of fields of activity. Polarity and emotions in language are especially relevant in the analysis of humour, since understanding the emotional patterns associated with humour can be useful in the development of artificial intelligence systems capable of detecting, understanding and generating humour effectively. The present study analyses data from the COLUMNAS HUMOR corpus (full texts and humorous sequences) with the aim of observing the polarity and predominant emotions therein. The methodology is corpus-based, with the use of digital tools to extract data on polarity and emotions and the application of quantitative and qualitative analysis. Drawing on different types of frequency-based findings, a qualitative analysis is conducted to interpret the data in each corpus under study. This exhaustive analysis has allowed us to detect differences in sentiment and emotional verbalisation in each corpus, as well as to contrast the results by corpus and by the gender of the studied journalists.

Keywords: sentiment analysis, emotions, humor, gender.

Sumario: 1. Introducción. Análisis de sentimiento, emociones y humor 2. La expresión verbal de las emociones relacionadas con el género. 3. Objetivos e hipótesis. 4. Material y métodos. 4.1. Corpus. 4.2. Procesamiento. 4.3. Convergencia de métodos. 5. Análisis y resultados. 5.1 Análisis de la polaridad y la intensidad. 5.2 Análisis de emociones. 6. Conclusiones. Agradecimientos. Referencias bibliográficas.

Cómo citar: Vargas-Sierra, C. (2023). Análisis emocional del corpus COLUMNAS.HUMOR: un enfoque mixto, *Círculo de Lingüística Aplicada a la Comunicación* 96, 73-87. <https://dx.doi.org/10.5209/clac.91592>

1. Introducción. Análisis de sentimiento, emociones y humor

Existe en la actualidad un claro interés investigador por analizar factores relacionados con las características cognitivas y fenómenos sociales a través del lenguaje, motivado, especialmente, por la posibilidad de acceder a infinidad

¹ Instituto Interuniversitario de Lenguas Modernas Aplicadas, Universidad de Alicante.
Correo electrónico: chelo.vargas@ua.es
ORCID: <http://orcid.org/0000-0002-4026-4372>

de textos disponibles en línea y generados por usuarios de todo tipo que interactúan en foros, blogs, redes sociales, noticias, etc. con distintas finalidades (dar su opinión, quejarse de la experiencia, de un producto, evaluarlo, etc.). Así, campos del procesamiento del lenguaje natural (PLN) como el análisis de sentimiento (AS), también llamado “minería de opinión”, y la detección computacional del humor representan solo una muestra de dicho interés.

La identificación de sentimientos y emociones en textos es una técnica que ha cobrado importancia en la investigación en los últimos años, ya que nos permite comprender mejor lo que las personas sienten y sus opiniones sobre diferentes temas. Esta línea permite analizar los aspectos afectivos del lenguaje, es decir, aquellos que van más allá de la mera información y transmiten emociones, sentimientos, actitudes y estados de ánimo. Detectarlos en los textos es importante para diversos campos, que van desde el análisis de la opinión de usuarios y clientes, hasta la detección de trastornos emocionales (depresión, ansiedad, bipolaridad, etc.) (Banchs, 2017). En definitiva, el análisis de sentimiento nos permite interpretar y comprender mejor la experiencia humana en diferentes contextos (Taboada, 2016).

En este marco, el estudio del humor y su relación con las emociones y la polaridad (positiva o negativa) en el lenguaje ha sido también un tema de interés creciente en la investigación en el Procesamiento de Lenguaje Natural (PLN) (Reyes et al., 2009b; Zhang et al., 2018; Mahajan y Zavery, 2020, entre otros), pues el análisis del significado del humor requiere una comprensión adecuada de las emociones y estados de ánimo que transmite el texto, por lo que resulta necesario ir más allá de lo que significa literalmente el lenguaje y considerar aspectos afectivos en el análisis del humor (Reyes et al., 2009b).

Todo el mundo sabe qué es lo que le hace reír o le pone alegre o ambas cosas y, sin embargo, desde los tiempos clásicos, se ha intentado definir el humor, aunque parece que sin éxito, pues no existe una teoría universalmente aceptada que explique lo que puede provocar la risa o lo que puede llevarnos a sentir que algo es divertido (Ruiz-Gurillo, 2019). Son múltiples las definiciones que podemos encontrar de humor y también las teorías desarrolladas para explicarlo, elaboradas por disciplinas como la filosofía, la psicología o la antropología (Attardo, 1994). Algunas son amplias y genéricas, como la de Raskin (1985: 46): “verbal humour is any text which is capable of creating a humorous effect”. De esta definición podemos destacar su inclusividad, dado que no limita el humor verbal a ningún tipo específico de texto o recurso lingüístico. Tampoco profundiza en los mecanismos específicos que crean el humor en el texto, sino que se enfoca en el efecto final, por lo que permite que diferentes perspectivas y enfoques teóricos puedan analizar y comprender el humor de forma variada e interdisciplinar. Por último, no distingue entre tipos o estilos de humor. A pesar de sus bondades al disponer de un elevado margen de aplicabilidad, contar con definiciones del humor más precisas y no abstractas en términos lingüísticos resulta ser crucial para poder detectarlo computacionalmente (Reyes, 2013: 108).

Además de la dificultad que entraña definirlo de forma precisa, también resulta ser una de las manifestaciones humanas más difíciles de aprehender “pues entran en juego aspectos tan diversos como el contexto, el género discursivo, los efectos que persigue el hablante/escritor o la comprensión adecuada que ha de hacer el oyente/lector” (Ruiz-Gurillo, 2019: 1231). Como vemos, el humor es una característica compleja y multifacética de la comunicación humana que involucra factores lingüísticos, culturales, cognitivos y psicológicos. Esta combinación de factores produce diversos tipos de humor y cada tipo tiene unas características específicas que lo diferencian de los demás, pues no es lo mismo el humor absurdo, que el humor negro, la sátira o la parodia. Es un fenómeno, por tanto, que no se puede explicar de forma aislada y que implica el uso de diversos y variados recursos de conocimiento lingüístico, narrativo, cultural, contextual, etc., lo que lo convierte en una dificultad a la hora de ser formalizado computacionalmente. Con todo, se han observado ciertas propiedades recurrentes, como la presencia de antónimos, de información sexual o de argot adulto (Reyes y Rosso, 2009: 101). Además, se ha descubierto que la ambigüedad semántica es un desencadenante común de efectos humorísticos y se ha reconocido la importancia de los rasgos afectivos, como emociones, sentimientos, actitudes y estados de ánimo (*ibid.*).

La detección computacional del humor es un campo en desarrollo que se enfrenta a numerosos desafíos y limitaciones (Cowie, 2023). Uno de ellos es la naturaleza subjetiva del humor, lo que dificulta la creación de algoritmos y modelos precisos para detectarlo en el texto. Además, el humor puede depender de cuestiones contextuales o culturales, lo que significa que lo que es divertido en un contexto o cultura puede no serlo en el otro. Otra limitación es la complejidad y variedad de los recursos lingüísticos utilizados en el humor verbal, como el sarcasmo, la ironía, el doble sentido, entre otros. Cada uno de estos recursos lingüísticos puede ser difícil de detectar automáticamente y requiere técnicas específicas de PLN.

Se han realizado numerosas investigaciones que se basan en el análisis de sentimiento y emociones, cuyas técnicas se utilizan para identificar patrones en los textos relacionados con la polaridad emocional y las emociones específicas que se evocan en el discurso (Orts, 2021; Nandwani y Verma, 2021; López-Rodríguez, 2022; Vargas-Sierra y Orts, 2023). Simplificando un concepto hartamente complejo y debatido por varias disciplinas, la emoción puede definirse como una respuesta mental específica, como el miedo, la alegría, la ira o el amor, que surge como reacción a estímulos internos o externos. Estas respuestas emocionales provocan cambios en nuestro cuerpo, expresiones faciales y comportamientos; así, el análisis de emociones determina exactamente qué estados emocionales o mentales se evidencian en un texto mediante la identificación de términos o frases referidos a la esfera afectiva y emocional implicada. Estas características, polaridad y emociones, pueden ser útiles para comprender mejor cómo se manifiesta el humor en el lenguaje y cómo se puede identificar en los textos, en muchas ocasiones a través de una señal de incongruencia. Dicho de otro modo, la aparición de palabras con polaridad opuesta o incongruentes en un mismo contexto puede ser una señal de humor y una característica importante para su detección (Reyes et al., 2009b; van den Beukel y Lora, 2018; Zhang et al., 2018; Mahajan y Zavery, 2020, entre otros). Las características basadas

en la emoción, por su parte, buscan identificar patrones humorísticos en textos a través de elementos que representan actitudes, sentimientos y estados de ánimo. En este contexto, nuestro estudio se enmarca en las emociones propuestas por Plutchik (2001), que identifica ocho emociones básicas: alegría, confianza, miedo, sorpresa, tristeza, ira, aversión y anticipación, que nos proporcionan un marco teórico para comprender cómo se expresan y se relacionan las diferentes dimensiones emocionales en los corpus que hemos estudiado.

2. La expresión verbal de las emociones relacionadas con el género

Es un estereotipo comúnmente aceptado que las mujeres son más emocionales que los hombres (Brody y Hall, 1993: 396; Barret et al., 1998; Cameron, 2005). Se afirma que la feminidad es más emocional y expresa con mayor intensidad los sentimientos, mientras que la masculinidad se asocia con la racionalidad y el autocontrol emocional (Fisher, 1993: 303; Fischer y Manstead, 2000: 71; Ahmadi-Azad, 2015: 620). Fisher (ibid.: 306) parte de la premisa de que tanto hombres como mujeres expresan la misma variedad de emociones y que las diferencias se deben, por un lado, a factores sociales y culturales, como el “omnipresent stereotype of the emotional woman or gender-specific feeling rules” y, por el otro, a que las mujeres las expresan más libremente.

Con respecto a la expresión verbal de las emociones, son numerosos los estudios que indican que es posible obtener una gran cantidad de información de estas a través del recuento y la categorización de las palabras que utilizan en su comunicación (Ahmadi-Azad, 2015; Bamman et al., 2012; Goldshmidt y Weller, 2000; Vainik, 2006; entre otros). En español, el vocabulario de las emociones consiste en sustantivos abstractos que denotan conceptos (‘miedo’, ‘felicidad’), en verbos (‘disfrutar’, ‘calmar’), algunos adverbios (‘maravillosamente’), adjetivos (‘nauseabundo’, ‘ilustre’), interjecciones (‘¡ay!’), ‘¡vaya!’), ‘¡uf!’) y participios adjetivales que denotan estados emocionales (‘abatido’, ‘eufórico’, ‘melancólico’).

En relación con la expresión emocional y el género, diversos estudios han abordado sus diferencias y, aunque no son significativas, sobresalen algunos patrones. Gilligan (2009: 18) afirma que las mujeres tienden a expresar más frecuentemente emociones como el amor, el miedo y la tristeza, mientras que los hombres son más propensos a expresar la ira. Bamman et al. (2012: 15-16) obtuvieron en el análisis de sus datos que las mujeres eran más propensas a utilizar palabras emocionales, emoticonos y jerga de la comunicación digital, como *lol* y *omg*. Los hombres, en cambio, utilizaban más números, palabras sobre tecnología y palabrotas. Vainik (2006), por su parte, encontró diferencias basadas en el género (también en la edad) en la forma en que las personas procesan y priorizan el lenguaje emocional. Los resultados mostraron que las mujeres tenían un conocimiento semántico de las emociones más accesible y diferenciado que su conocimiento episódico. Por “conocimiento semántico” se hace alusión al conocimiento general y abstracto de las palabras y su significado, mientras que el “conocimiento episódico” se refiere a las experiencias personales y específicas asociadas con esas palabras. Además, había una mayor diferencia en la forma en que los hombres y mujeres procesaban los términos básicos y no básicos de las emociones. Goldshmidt y Weller (2000) exploraron las diferencias de género en el lenguaje y las emociones, consideradas como derivadas de la diferente socialización de los sexos. En concreto, examinaron si las mujeres utilizaban más palabras emocionales que los hombres, con el resultado de que en todas las muestras observadas hallaron que las palabras emocionales aparecían con más frecuencia en la comunicación femenina. Chaplin y Aldao (2013), por su parte, estudiaron las diferencias de género en la expresión de emociones desde la infancia hasta la adolescencia, y encontraron pequeñas pero significativas diferencias, pues las niñas mostraban más frecuentemente emociones positivas e internas, como la tristeza, la ansiedad, la compasión, que los niños, quienes, por su parte, mostraban más emociones externas, sobre todo la ira. En particular, las diferencias de género se vieron moderadas por la edad, el contexto interpersonal y el tono emocional de la tarea que se daba a los participantes, lo que subraya la importancia de los factores contextuales en las diferencias de género.

Ahmadi-Azad (2015) examinó el contenido emocional de informes autobiográficos escritos por 103 mujeres y 82 hombres estudiantes de inglés como lengua extranjera. Se utilizó un análisis de procesamiento del lenguaje en línea para determinar el contenido emocional lingüístico de los datos recopilados. Aunque el análisis estadístico no mostró diferencias significativas entre los géneros, sí indicaban que las mujeres tienden a utilizar un mayor número de palabras emocionales que los hombres.

En consecuencia, según estos estudios apuntados, las mujeres suelen ser más expresivas emocionalmente que los hombres, es decir, manifiestan una mayor intensidad emocional. Por otra parte, los hombres utilizarían más palabras emocionales negativas que las mujeres, es decir, con un tono emocional tendente a la negatividad.

En nuestro estudio clasificaremos el sentimiento y las emociones con el objetivo de comprender mejor el modo en que se manifiesta el humor en los corpus estudiados y si es posible identificarlo a través de patrones emocionales y de polaridad. Además, desvelaremos las características emocionales en una serie de artículos de opinión y en sus secuencias de humor con el fin de proporcionar una perspectiva más detallada y específica sobre cómo las emociones y la polaridad se relacionan con el humor y en diferentes autores.

3. Objetivos e hipótesis

Partiendo de la premisa de que el análisis del sentimiento y las emociones se relacionan con la expresión del humor, la presente investigación tiene como objetivos los siguientes:

- (1) O1: identificar aspectos afectivos que puedan ayudar a la tipificación de patrones humorísticos contrastando los artículos de opinión con sus secuencias humorísticas.
- (2) O2: conocer mejor la polaridad y las emociones expresadas en los textos de los/as periodistas analizados contrastando sus corpus.
- (3) O3: detectar las diferencias en la expresión emocional (sentimiento y emociones) entre los autores y las autoras estudiados en los diferentes corpus creados al efecto.

Las hipótesis de partida son:

- (1) H1: Los corpus de secuencias humorísticas contienen sentimientos más intensos y predominantes de carácter positivo, en comparación con el corpus que contiene los artículos de opinión completos.
- (2) H2: Existe una mayor presencia de emociones negativas, como la tristeza y la ira en los textos completos a causa del tipo de texto que se trata, el artículo de opinión, a diferencia de lo que ocurrirá en el corpus de secuencias humorísticas, donde el humor se utiliza para suavizar el tono general del artículo.
- (3) H3: Los autores y autoras presentan en sus textos una polaridad e intensidad similar, y lo mismo ocurre con sus emociones, pero con una posible mayor presencia de polaridad y emociones positivas en los textos de las autoras.

Se utilizará una convergencia de métodos, es decir, combinaremos diferentes técnicas y enfoques metodológicos para realizar un análisis del discurso con el objetivo de demostrar las tres hipótesis anteriores. Más concretamente, nos servimos del análisis de la polaridad y de las emociones, que nos permitirá medir el grado de polarización de los corpus a partir de los sentimientos dominantes. Por otro lado, realizaremos el análisis de las emociones, que hará posible analizar el léxico empleado en la elaboración de los artículos de opinión analizados y de sus secuencias humorísticas y relacionarlo con el género de los autores.

En la siguiente sección de este trabajo, se abordará la metodología utilizada para el análisis realizado. Se describirán los corpus estudiados, el proceso utilizado para recopilar y prepararlos para su análisis, así como las herramientas y técnicas utilizadas para la obtención de datos.

4. Material y métodos

4.1. Corpus

Para lograr los objetivos antes apuntados, se compilaron y se procesaron los textos completos del corpus COLUMNAS.HUMOR², que se compone de 130 secuencias humorísticas³, definidas como enunciados que contienen marcas e identificadores humorísticos (Ortega-Gilabert y Timofeeva-Timofeev, 2023). El corpus completo contiene 83 artículos de opinión de seis reputados y reputadas columnistas publicados en diarios españoles de prestigio. Los autores/as son: Ignacio Escolar (IE), Moncho Alpuente (MA), Juan J. Millás (JJM), Rosa Montero (RM), Carmen Rigalt (CR) y Elvira Lindo (EL).

Primero se trabajó con el corpus completo y se extrajeron datos de polaridad y emociones para cada autor/a. Luego, se contrastaron estos datos con los obtenidos del corpus que solo contenía segmentos humorísticos; ambos fueron procesados utilizando herramientas informáticas, como explicamos en la siguiente sección, para obtener datos cuantitativos y cualitativos sobre polaridad, intensidad y emociones. A partir de estos resultados, se realizó un análisis cuantitativo para describir las diferencias en el sentimiento y la expresión emocional en cada corpus, y un análisis cualitativo para hacer interpretaciones y análisis más detallados en cada corpus de estudio.

Cabe destacar que la naturaleza lingüística y emocional de los artículos de opinión que conforman el corpus COLUMNAS.HUMOR, a diferencia de los textos estrictamente humorísticos, presenta una complejidad mayor en su análisis. En este tipo de textos, el humor no es el objetivo principal, sino que es un recurso utilizado para reforzar la argumentación o para aligerar el tono de la opinión expresada. Por lo tanto, el análisis de los ítems emocionales presentes en los corpus se llevó a cabo con el objetivo de estudiar las emociones y el sentimiento que se expresan en ellos, y si presentan diferencias al comparar el corpus completo con el de las secuencias por autores.

La siguiente tabla (Tabla 1) presenta los datos de tamaño del corpus textual y el de las secuencias, así como de sus subcorpus (autores hombres o mujeres):

² Puede obtenerse una descripción completa del corpus en Martínez-Egido (2014).

³ Disponibles para su consulta en: Observa Humor (<http://www.observahumor.com/corpus>).

	Archivos de corpus A1	Palabras	Archivos de corpus B1	Palabras
Subcorpus autores	H_MA_completo.txt	9.625	H_SecuMA.txt	2.791
	H_JJM_completo.txt	5.758	H_SecuJJM.txt	1.031
	H_IE_completo.txt	5.291	H_SecuIE.txt	485
	TOTAL	20.674		4.307
	Archivos de corpus A2		Archivos de corpus B2	
Subcorpus autoras	M_RM_completo.txt	7.339	M_SecuRM.txt	2.008
	M_CR_completo.txt	6.634	M_SecuCR.txt	1.684
	M_EL_completo.txt	5.558	M_SecuEl.txt	796
	TOTAL	19.531		4.488
TOTAL CORPUS		40.205		8.795

Tabla 1. Composición del corpus y de cada subcorpus

Las cifras de la Tabla 1 muestran que el corpus completo está compuesto por un total de 40.205 palabras y su subcorpus de secuencias humorísticas⁴ contiene 8.795 palabras. Es decir que el 21,89 % del corpus se identificó como secuencias humorísticas, lo que sugiere que el humor es un tema recurrente en los textos analizados, aunque no sea el tema principal. El subcorpus A1 cuenta con un total de 20.674 palabras y 4.307 palabras corresponden a su corpus de secuencias humorísticas (B1), lo que equivale al 20,83 % del subcorpus. Esto apunta a que los autores tienen una proporción similar de secuencias de humor en sus textos en comparación con el corpus completo. El subcorpus de autoras (A2) cuenta con un total de 19.531 palabras y 4.488 palabras corresponden a su corpus de secuencias humorísticas (B2), lo que equivale al 22,96 % del subcorpus, por lo que deducimos que las autoras estudiadas tienden a utilizar ligeramente más el humor en sus textos en comparación con los autores. Estos datos indican que el humor es un tema recurrente en los textos analizados y que hay una distribución equilibrada de palabras entre los subcorpus de autores y autoras, aunque estas últimas tienden a utilizar un poco más el humor en sus textos en comparación con los autores.

4.2. Procesamiento

En la siguiente imagen (Figura 1) obtenemos una síntesis de lo que ha implicado el proceso para llevar a cabo el estudio que nos ocupa:

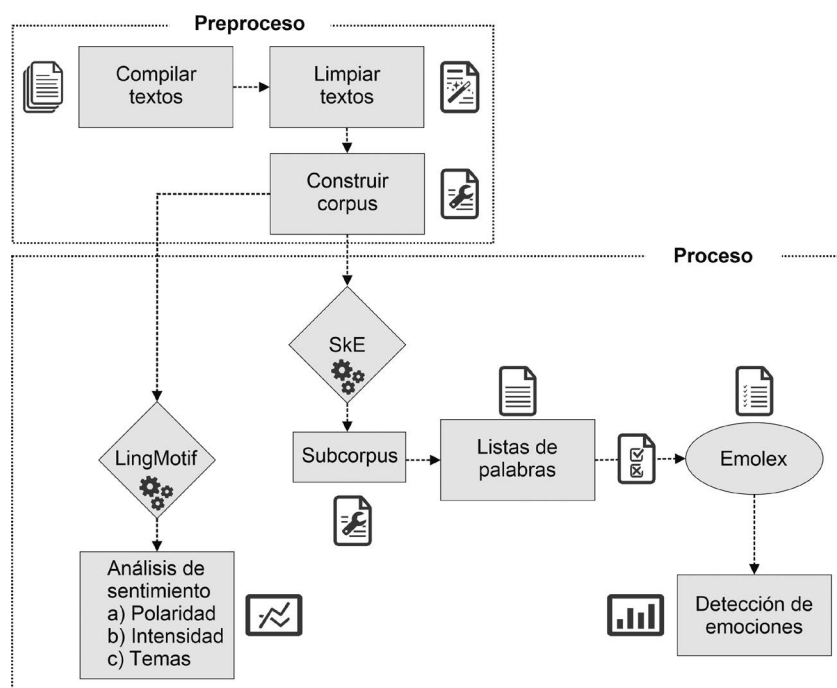


Figura 1. Proceso de constitución de corpus y procesamiento (adaptada de Vargas-Sierra y Orts, 2023)

⁴ Sobre el modo de detectar y etiquetar las secuencias humorísticas en los textos véase Ortega-Gilbert y Timofeeva-Timofeev (2023).

Como se aprecia en la anterior figura, la metodología se compone de varias fases, a saber:

- (1) Fase 1, preproceso: En esta primera fase se prepara el corpus COLUMNAS.HUMOR para su análisis posterior. Esto implica la compilación de los textos completos y de las secuencias humorísticas (desde la plataforma OBServa.HUMOR), así como su limpieza y organización. Aquí se crearon también los subcorpus según el género de los periodistas.
- (2) Fase 2, proceso con Lingmotif: En esta segunda fase se utiliza esta herramienta informática para procesar el corpus y obtener información sobre la polaridad (sentimientos positivos o negativos) y la intensidad emocional de los textos. Lingmotif también permite identificar los temas más recurrentes en el corpus analizado.
- (3) Fase 3, proceso con Sketch Engine: En esta tercera fase se utiliza este programa de análisis textual para procesar el corpus y obtener información más detallada sobre las palabras y expresiones utilizadas en el corpus. Esto incluye la creación de listados de palabras más frecuentes en cada subcorpus según su categoría gramatical.
- (4) Fase 4, contraste con EmoLex: En esta última fase se contrastan automáticamente los listados de palabras del lexicón NRC Emotion Lexicon, consistente en un listado de palabras asociadas a ocho emociones básicas (Plutchik, 2001; Plutchik y Kellerman 2013), con los listados de frecuencia por categoría gramatical. Este contraste permite detectar las emociones más frecuentes en cada subcorpus y analizar su relación con la polaridad de los textos.

4.3. Convergencia de métodos

Para lograr los objetivos de la investigación, se han utilizado varias técnicas metodológicas. Por un lado, se ha llevado a cabo un análisis de sentimiento para extraer datos de polaridad e intensidad. Por el otro, se ha realizado un análisis de léxico emocional recurrente para estudiar las emociones presentes en los textos cuantitativa y cualitativamente del corpus COLUMNAS.HUMOR. Estas técnicas se distribuyeron en distintas fases (Figura 1).

En la fase 1 analizamos la polaridad y la intensidad de los corpus A1, A2, B1 y B2 (Tabla 1) y para ello se utilizó Lingmotif 2 (Moreno-Ortiz, 2021). Se trata de una herramienta de análisis textual pero dirigida al AS. Utiliza un enfoque basado en diccionarios para detectar la positividad o negatividad contenida en las expresiones lingüísticas de uno o varios textos; es decir, compara las unidades léxicas del corpus de entrada (*input*) con las contenidas en su propio diccionario interno, que incluye palabras y frases con carga afectiva previamente etiquetadas como positivas o negativas. Lingmotif utiliza una escala de 0 a 100 para categorizar los textos según su polaridad, que va desde ‘extremadamente negativo’ (-100) a ‘extremadamente positivo’ (+100), o 0 para la neutralidad; todo ello, basándose en la orientación semántica del sentimiento detectado en el texto, esto es, mediante una evaluación del texto según transmita un sentimiento positivo, negativo o neutro. El *Text Sentiment Score* (TSS) es el valor que se asigna a la polaridad de cada oración en el texto, y para ello el software tiene en cuenta tanto la cantidad de elementos con carga afectiva, como sus respectivas posiciones en la oración. Calcula también la intensidad del sentimiento del texto (TSI) ponderando el número de oraciones positivas y negativas. El software utiliza reglas de contexto (llamadas CVS o *context valence system*) para acomodar posibles modificadores de sentimiento, como la inversión, la intensificación y la atenuación. A los elementos de sentimiento se le asignan valores de -5 a +5, donde 0 indica una expresión de neutralidad. De esta manera, Lingmotif puede proporcionar una medida cuantitativa de la polaridad del texto y la intensidad del sentimiento, lo que permite comparar y analizar los textos con mayor precisión (Moreno-Ortiz, 2017: 133).

Muestra dos tablas de resultados: *Sum and Averages* y *Document List*. La primera ofrece una serie de puntuaciones y elementos de sentimiento y características lingüísticas de todo el corpus y la segunda muestra, además, las categorías de sentimiento de cada uno de los textos. Por tanto, podemos obtener resultados globales o específicos como los siguientes:

- Puntuaciones y elementos de sentimiento: TSS, o puntuación total del sentimiento del documento; TSI, o puntuación total de la intensidad del sentimiento del documento; SA ITEMS, el número de ítems con sentimiento afectivo; POS ITEMS, el número de elementos positivos; NEG ITEMS, número de elementos negativos; NEU ITEMS, número de elementos neutros; POS SENTENCES, número de frases positivas; NEG SENTENCES, número de frases negativas; y NEU SENTENCES, número de frases neutras.
- Categorías de sentimiento: TSS CAT2 (basada en una escala de dos puntos, es decir, positivo y negativo; TSS CAT3 (basada en una escala de tres puntos, positivo, negativo y neutro; TSS CAT9 (basada en una escala de nueve puntos (de extremadamente negativo a extremadamente positivo); TSI CAT o categoría de intensidad del sentimiento basada en una escala de cinco puntos (de muy débil a muy fuerte).
- Características lingüísticas, como el número de palabras (*tokens*), la diversidad léxica (*type/token ratio*), las palabras léxicas o de contenido (*lex words*), las funcionales (*function words*), las palabras monoléxicas, poliléxicas (*multiwords*), el número de palabras/frases repetidas, mayúsculas, nombres propios, comunes, verbos, adjetivos, adverbios, puntuación, signos de interrogación (*q marks*), de exclamación (*x marks*), numerales, etiquetas (*hashtags*), menciones de usuarios (*handles*), URL y emojis.

En la fase 3, con Sketch Engine (Kilgarrif et al., 2004), obtuvimos los listados frecuenciales de sustantivos, verbos y adjetivos de los corpus, al ser las categorías más estrechamente relacionadas con la expresión y comunica-

ción de emociones en el lenguaje. Los adverbios pueden proporcionar información adicional sobre la intensidad o la forma de una expresión emocional, pero las descartamos en este estudio por cuestiones operativas, al ser una categoría que no siempre está directamente relacionada con la expresión de emoción y tener, más frecuentemente, la función de modificador de verbos o adjetivos, y así lo vimos en los listados.

WORDLIST

Hombres **adverb** (181 items | 1,021 total frequency)

	Adverb	Frequency [?] ↓	Frequency Per Million [?] ↓
1	no	268	11,547.74 ...
2	más	115	4,955.19 ...
3	tanto	45	1,938.99 ...
4	también	32	1,378.83 ...
5	menos	30	1,292.66 ...
6	ya	30	1,292.66 ...
7	muy	28	1,206.48 ...
8	bien	25	1,077.21 ...
9	aun	19	818.68 ...
10	solo	19	818.68 ...
11	siempre	18	775.59 ...
12	mucho	18	775.59 ...
13	ahí	18	775.59 ...

Figura 2. Listados de adverbios por frecuencia

Estas listas nos sirvieron para proceder con la siguiente fase, la 4, pues las contrastamos con un lexicón, denominado *NRC Word-Emotion Association Lexicon*, también conocido como EmoLex (Mohammad y Turney 2013; Mohammad, 2018; Mohammad, 2020). Este recurso contiene las ocho emociones básicas de Plutchik (2001), esto es, ALEGRÍA (emoción positiva asociada con sentimientos de felicidad, satisfacción y entusiasmo y que se experimenta en respuesta a eventos o situaciones que son percibidos como gratificantes o positivos); CONFIANZA (surge cuando se percibe seguridad, fiabilidad y credibilidad en uno mismo o en los demás, caracterizada por la creencia en la integridad y la capacidad de cumplir con las expectativas); MIEDO (emoción negativa relacionada con la percepción de amenaza o peligro que se experimenta como una respuesta de alerta y preparación para enfrentar situaciones potencialmente dañinas); ANTICIPACIÓN (surge cuando se espera o se anticipa algo en el futuro y puede manifestarse como emoción positiva, como el entusiasmo, o emoción negativa, como la ansiedad, dependiendo del contexto y las expectativas asociadas); TRISTEZA (emoción negativa asociada con sentimientos de pesar, desánimo y pérdida, que se experimenta en respuesta a situaciones que implican separación, decepción o desgaste emocional), AVERSIÓN (surge como respuesta a estímulos o situaciones que se consideran desagradables o repulsivos, que se identifica por el rechazo y la repulsión hacia lo que se percibe como negativo o perjudicial), IRA (emoción negativa intensa relacionada con la frustración, la irritación y la rabia, que asoma como respuesta a la percepción de injusticia, violación de expectativas o amenazas percibidas); y, por último, SORPRESA (aparece al experimentar algo inesperado o novedoso y puede ser tanto positiva como negativa). Además de seguir el modelo emocional de Plutchik (*ibid.*), el lexicón contiene dos dimensiones de polaridad (positiva y negativa) y se compone de más de 14.000 unigramas en inglés y sus equivalentes en otros idiomas, incluido el español, cuya última traducción es del 2022⁵. El EmoLex hace una categorización binaria de las palabras en positivas o negativas y en una o más emociones, ya que no son mutuamente excluyentes. Cada palabra en el léxico se evalúa en términos de su carga emocional en positivo o negativo y en alguna o varias emociones con la puntuación 0 o 1 (Figura 3), lo que indica el grado de asociación de la palabra con la positividad/negatividad y la emoción o emociones concretas:

⁵ Las versiones de EmoLex a otras lenguas, 108 en total, se han realizado utilizando el traductor automático de Google. La última actualización es de agosto de 2022, según se indica en la página del autor: <https://saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm>

Spanish Word	negative	positive	anger	anticipation	disgust	fear	joy	sadness	surprise	trust
avance	0	1	0	0	0	0	0	0	0	0
avanzando	0	0	0	0	0	0	0	0	0	0
ventaja	0	1	0	0	0	0	0	0	0	0
ventajoso	0	1	0	0	0	0	0	0	0	0
adviento	0	1	0	1	0	0	1	0	0	1
aventura	0	1	0	1	0	0	0	0	0	0
aventurero	0	0	0	0	0	0	0	0	0	0
aventurero	0	1	0	0	0	0	0	0	0	0
adversario	1	0	1	0	0	0	0	0	0	0
adverso	1	0	1	0	1	1	0	1	0	0
adversidad	1	0	1	0	0	1	0	1	0	0
anunciar	0	0	0	0	0	0	0	0	0	0
anuncio publicit	0	0	0	0	0	0	0	0	0	0
consejo	0	0	0	0	0	0	0	0	0	1
aconsejable	0	1	0	0	0	0	0	0	0	1

Figura 3. Muestra de EmoLex en español

La asignación de emociones a una palabra en nuestras listas de frecuencia se realizó automáticamente mediante fórmulas avanzadas de búsqueda condicional de Excel, como ya explicamos en Vargas-Sierra y Orts (2023). El resultado de la función de búsqueda automática es una lista de unigramas por frecuencia con la polaridad y las emociones marcadas con el valor 1, como se muestra en la siguiente figura, en la que se asignaron diferentes colores para facilitar la identificación:

Item	Frequency	Relative frequenc	Positive	Negative	Anger	Anticipation	Disgust	Fear	Joy	Sadness	Surprise	Trust
avergonzar	1	44.87726	0	1	0	0	0	0	0	0	1	0
llorar	1	44.87726	0	1	0	0	0	0	0	0	1	0
sospechar	2	89.75452	0	1	0	0	0	1	0	0	0	0
juzgar	2	89.75452	0	1	0	0	0	1	0	0	0	0
castigar	1	44.87726	0	1	0	0	0	1	0	0	0	0
ahogar	3	134.63178	0	1	0	0	0	1	0	1	0	0
deprimir	1	44.87726	0	1	0	0	0	1	0	1	0	0
matar	1	44.87726	0	1	0	0	0	1	0	1	0	0
asustar	1	44.87726	0	1	0	0	0	1	0	1	1	0
infectar	3	134.63178	0	1	0	0	1	0	0	0	0	0
contaminar	1	44.87726	0	1	0	0	1	0	0	0	0	0

Figura 4. Muestra de resultado del contraste de una lista de verbos con EmoLex

5. Análisis y resultados

A partir de los resultados obtenidos en el preproceso, proceso del corpus y contraste con el lexicón, se llevó a cabo el análisis cuantitativo de los datos, extraídos por parámetros de frecuencia, polaridad e intensidad. Una vez obtenidos dichos datos cuantitativos, se procedió a realizar el análisis cualitativo, que nos permitió hacer interpretaciones en cada corpus de estudio. Se examinaron las palabras más utilizadas y las emociones predominantes para comprender la naturaleza y el contenido de los textos y cómo se expresan los sentimientos y emociones en ellos. Como se verá, este análisis nos proporcionó una descripción bastante exhaustiva de las diferencias en el sentimiento y la verbalización emocional en cada corpus, lo que nos permitió conocer la forma en que los periodistas utilizaban el humor y cómo expresaban sus emociones y sentimientos en sus textos.

5.1. Análisis de la polaridad y la intensidad

La primera aproximación se centró en la identificación de la polaridad y la intensidad de los corpus A1 y A2 para determinar la orientación perfilada en estos dos corpus.

La siguiente figura muestra el promedio de la puntuación global del sentimiento del texto (TSS por sus siglas en inglés) y el promedio de la intensidad en el texto (TSI) de los corpus A1 y A2.

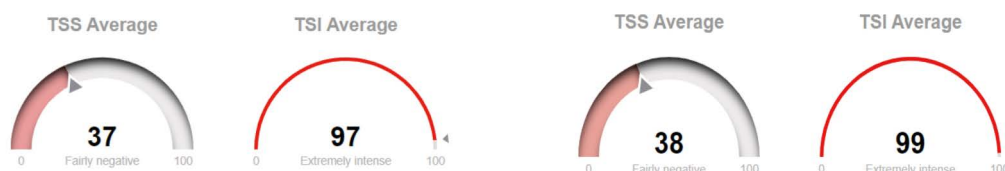


Figura 5. Resultados de sentimiento e intensidad de los corpus A1 (izquierda) y A2 (derecha)

Como se aprecia en el indicador gráfico (Figura 5), la puntuación global de sentimiento o tss, tanto en el corpus de autores (37, *fairly negative*), como en el de autoras (38, *fairly negative*) es bastante negativa, ligeramente mayor en las autoras, y la intensidad, es decir, la proporción de elementos de sentimiento frente a los que no lo son, es extremadamente intensa (*extremely intense*) en ambos corpus, también levemente superior en el corpus A2 (99 frente a 97).

Además de estas puntuaciones, que son los primeros datos que muestra gráficamente Lingmotif, este programa, en la tabla *Sums and Averages*, ofrece diversas características. Por cuestiones de objetivos planteados y espacio, ofrecemos exclusivamente los resultados obtenidos que están relacionados con el sentimiento para los dos corpus (Tabla 2):

Corpus A1							
TSS		TSI		SA ITEMS		POS ITEMS	
SUM	AVG	SUM	AVG	SUM	AVG	SUM	AVG
110	36,67	290	96,67	10275	3425	634	211,33
NEU ITEMS		POS SENTENCES		NEG SENTENCES		NEU SENTENCES	
SUM	AVG	SUM	AVG	SUM	AVG	SUM	AVG
422	140,67	254	84,67	373	124,33	256	85,33
Corpus A2							
TSS		TSI		SA ITEMS		POS ITEMS	
SUM	AVG	SUM	AVG	SUM	AVG	SUM	AVG
115	38,33	296	98,67	9980	3326,67	581	193,67
NEU ITEMS		POS SENTENCES		NEG SENTENCES		NEU SENTENCES	
SUM	AVG	SUM	AVG	SUM	AVG	SUM	AVG
442	147,33	294	98	381	127	386	128,67

Tabla 2. *Sums and Averages* del corpus A1 y A2

Además del TSS y TSI, ya comentados más arriba, en la Tabla 2 encontramos el número total de ítems (*items*) y oraciones (*sentences*) de los dos corpus, clasificados como positivos, negativos, neutros o con un sentimiento afectivo en particular (SA). Estos datos sugieren que el corpus A1 tiende a tener un tono bastante negativo, ya que el promedio TSS es de 38,33, que está en el rango de *fairly negative* según la escala de Lingmotif. Además, el promedio de TSI es extremadamente alto, lo que indica que el texto es intenso en términos de polaridad emocional. La Tabla 2 muestra también que hay un número significativo de SA ITEMS (10,275), lo que sugiere que hay un uso frecuente de palabras con connotaciones negativas en el corpus. Además, el número de elementos positivos (POS ITEMS, 634) es relativamente bajo en comparación con los SA ITEMS, lo que indica que hay una menor frecuencia de palabras con connotaciones positivas en el corpus.

En cuanto a la distribución de sentimientos en las distintas categorías, la columna de NEU ITEMS (ítems de sentimiento neutro) tiene un total de 422, con un promedio de 140,67, lo que sugiere que los textos tienen un grado moderado de neutralidad en cuanto a sentimientos. Además, en la columna de POS SENTENCES (oraciones con sentimiento positivo) el corpus tiene un total de 254, con un promedio de 84,67, lo que indica que las oraciones positivas son menos frecuentes que las neutras o negativas. En la columna de NEG SENTENCES (oraciones con sentimiento negativo) el corpus tiene un total de 373, con un promedio de 124,33, por lo que las oraciones negativas son más frecuentes que las positivas. En la columna de NEU SENTENCES (oraciones con sentimiento neutro), el corpus tiene un total de 256, con un promedio de 85,33, indicativo de que las oraciones neutras son relativamente frecuentes en los textos analizados. Estos resultados sugieren que este corpus tiene un alto grado de intensidad de sentimiento, pero que la polaridad predominante es negativa.

Con respecto al corpus A2, observamos que los valores son bastante similares a los del corpus de humor de los periodistas hombres (A1), aunque con algunas pequeñas diferencias. Por ejemplo, el promedio de TSS es ligeramente superior (38,33) y el de TSI es prácticamente igual (98,67). Los datos sobre otras características de polaridad del corpus A2 (Tabla 2) muestran que hay un número significativo de SA ITEMS (9.980), con un promedio de 3.326,67, ligeramente menor que en el corpus A1, y sugiere, igualmente, que hay un uso frecuente de palabras con connotaciones negativas en el corpus. Además, el número de elementos positivos (POS ITEMS, 581) y su promedio (193,67) es relativamente bajo en comparación con los SA ITEMS, lo que indica que hay una menor frecuencia de palabras con connotaciones positivas en el corpus. En comparación con el corpus A1, este promedio es menor (193,67 frente a 211). En cuanto a las características de las oraciones, el corpus A2 contiene un total de 442 palabras con polaridad neutra, con un promedio de 147,33. Además, hay un total de 294 oraciones con polaridad positiva y un promedio de 98, mientras que hay un total de 381 negativas y un promedio de 127. Estos datos apuntan a que este corpus de autoras tiene una presencia ligeramente mayor de emociones positivas y también un mayor número de palabras con

polaridad negativa. Ambos corpus presentan una fuerte presencia de emociones en los textos, lo que sugiere que las opiniones expresadas en ambos son altamente subjetivas y emotivas, como corresponde con el tipo de texto que estamos tratando.

En resumen, al comparar ahora los dos corpus, apreciamos que tanto los autores como las autoras evocan un sentimiento negativo en sus artículos, con una intensidad extremadamente alta en ambos casos, si bien ligeramente superior en el caso de las autoras. Se observa que los autores presentan más SA ITEMS que las autoras, lo que indica una mayor actividad emocional en sus artículos. Ellas despliegan un mayor número de oraciones negativas, lo que sugiere una tendencia hacia opiniones más críticas o negativas en sus escritos y, en contrapartida, los periodistas masculinos tienen un número ligeramente superior de oraciones positivas, pero un menor número de ítems y oraciones neutras, lo que puede indicar un enfoque menos equilibrado y matizado de sus temas.

La segunda aproximación la realizamos cargando en Lingmotif los corpus B1 y B2, es decir, los que solo contienen las secuencias humorísticas y obtuvimos los siguientes indicadores gráficos de los promedios TSS y TSI (Figura 6):

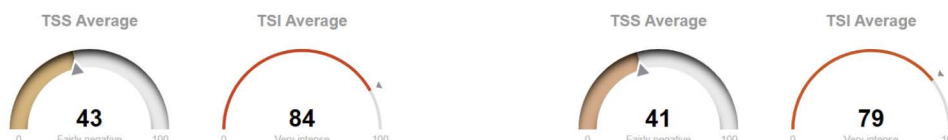


Figura 6. Resultados de sentimiento e intensidad de los corpus B1 (izquierda) y B2 (derecha)

Como se aprecia en la anterior figura, existe una ligera diferencia entre autores y autoras en lo que se refiere al promedio TSS y al TSI, lo que indica que las secuencias de humor de los autores pueden ser más elevadas e intensas en términos de sentimiento. La tabla siguiente expone con más detalle los elementos de sentimiento de los dos corpus (Tabla 3)

Corpus B1							
TSS		TSI		SA ITEMS		POS ITEMS	
SUM	AVG	SUM	AVG	SUM	AVG	SUM	AVG
130	43,33	251	83,67	2165	721,67	146	48,67
NEU ITEMS		POS SENTENCES		NEG SENTENCES		NEU SENTENCES	
SUM	AVG	SUM	AVG	SUM	AVG	SUM	AVG
84	28	58	19,33	61	20,33	40	13,33
Corpus B2							
TSS		TSI		SA ITEMS		POS ITEMS	
SUM	AVG	SUM	AVG	SUM	AVG	SUM	AVG
124	41,33	236	78,67	2307	769	134	44,67
NEU ITEMS		POS SENTENCES		NEG SENTENCES		NEU SENTENCES	
SUM	AVG	SUM	AVG	SUM	AVG	SUM	AVG
90	30	68	22,67	84	28	80	26,67

Tabla 3. Sums and Averages del corpus B1 y B2

Observamos que el corpus B1 tiene un menor número de SA ITEMS si lo comparamos con B2, lo que puede sugerir que las frases de humor escritas por los autores son menos emotivas que las escritas por las autoras. Si se comparan los resultados de los ítems POS y NEU, se observa que el corpus B1 tiene un mayor número de ítems positivos (146) que el B2, que tiene 134. Sin embargo, el corpus B2 tiene un mayor número de ítems neutros (90) que el masculino (84), lo que sugiere que las oraciones humorísticas del corpus B1 pueden tener un tono más positivo que las frases humorísticas del corpus B2. Si observamos los resultados de las oraciones (*sentences*) POS, NEG y NEU, vemos que el corpus B1 tiene menos oraciones positivas y negativas (58 y 61 respectivamente) que el corpus B2, con 68 y 84. Este último también tiene más frases neutras (80), el doble que el B1, que tiene 40. En resumen, el corpus B2 parece tener un enfoque más equilibrado en cuanto al tono de las frases humorísticas, con una mayor cantidad de ítems neutros y un número similar de oraciones positivas y negativas, mientras que el corpus B1 parece estar más enfocado en un tono favorable, con menos oraciones negativas y neutras.

Comparemos ahora los datos entre los corpus completos de los periodistas y los de sus secuencias humorísticas. Observamos que los valores promedios de TSS para el corpus B1 (43,33) es mayor que el mismo valor para el corpus A1 (36,67). Esto sugiere que el corpus de secuencias humorísticas contiene, según estos datos, más senti-

mientos polarizados en comparación con el corpus de artículos de opinión completos. Es decir, en el corpus B1 hay una mayor concentración de sentimientos fuertes y divergentes, lo que puede ser atribuible al hecho de que el humor utilizado ofrezca una expresión emocional exagerada y una polarización de los sentimientos. Por otro lado, el corpus A1 tiene una intensidad emocional (TSI) promedio más alta que el corpus de secuencias humorísticas.

A partir de los promedios de TSS y TSI de los corpus A2 y B2 se puede observar que ocurre lo mismo que con los autores, es decir, el promedio de los textos de B2 ofrecen un mayor TSS (41,33 vs 38,33) y un menor TSI (78,67) que los artículos de opinión contenidos en A2 (98,67).

Para determinar la razón de un TSI más elevado en los corpus A1 y A2 frente a los de secuencias humorísticas necesitaríamos evaluar internamente los textos, pero podría explicarse por el uso de palabras emocionales que tienen una polaridad más fuerte por el hecho de tratar temas que despiertan emociones más intensas, como corresponde al tipo textual. Por otro lado, en las secuencias humorísticas, los autores utilizan un lenguaje más ligero y entretenido, y esto puede ser la razón de la disminución de la puntuación del TSI. Los datos sugieren que las secuencias humorísticas tienen un mayor contenido emocional en general, pero una menor polarización en cuanto a la intensidad del sentimiento expresado. Esto podría ser indicativo de que aquí en el uso del humor se ha buscado generar una respuesta emocional en el lector o audiencia, pero sin llegar a extremos de polarización que puedan resultar desfavorables.

5.2. Análisis de emociones

Para el estudio de las emociones predominantes, cotejamos los nombres, adjetivos y verbos de los corpus A1, A2, B1 y B2 con el EmoLex. A continuación, presentamos los valores para cada corpus.

El primer gráfico (Figura 7) muestra los resultados al contrastar los corpus A1 y A2 en porcentajes:

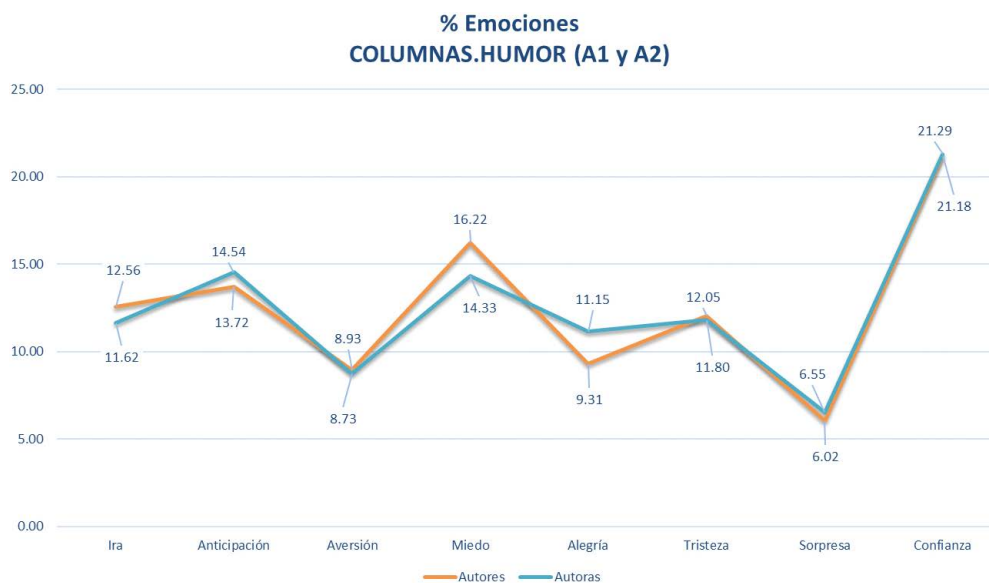


Figura 7. Emociones en los corpus A1 y A2 en porcentajes

Los valores representan el porcentaje de cada categoría de emoción en el lenguaje utilizado por autores y autoras, respectivamente. De acuerdo con los datos, la emoción más predominante en ambos grupos estudiados es la CONFIANZA, con un valor del 21,18 % y 21,29 %, respectivamente. En cuanto a las demás emociones, para los periodistas hombres la emoción más prevalente después de la confianza es el MIEDO (16,22 %), seguida de la ANTICIPACIÓN (13,72 %), la IRA (12,56 %), la TRISTEZA (12,05 %), la AVERSIÓN (8,93 %), la SORPRESA (6,02 %), y finalmente la ALEGRÍA (9,31 %). Para las periodistas mujeres, la emoción más prevalente después de la CONFIANZA es la ANTICIPACIÓN (14,54 %), seguida del MIEDO (14,33 %), la IRA (11,62 %), la TRISTEZA (11,80 %), la ALEGRÍA (11,15 %), la AVERSIÓN (8,73 %) y, finalmente, la SORPRESA (6,55 %).

Las diferencias que destacamos a partir de estos datos son las siguientes:

- La emoción predominante es la misma para ambos grupos (CONFIANZA), pero hay una pequeña diferencia en el porcentaje a favor de las autoras (21,18 % para los autores y 21,29 % para las autoras).
- En cuanto al resto de las emociones que aparecen, los periodistas presentan un porcentaje mayor de MIEDO (16,22 % vs 14,33 %), de TRISTEZA (12,05 % vs 11,8 %) de IRA (12,56 % vs 11,62 %), de AVERSIÓN (8,93 % vs 8,73 %) y de SORPRESA (6,55 % vs 6,02 %). De estas emociones, las diferencias más destacables se sitúan en el MIEDO y la IRA, casi con un punto por encima.
- Las periodistas, en cuanto al resto de emociones, tienen un porcentaje mayor de ANTICIPACIÓN (14,54 %) y de ALEGRÍA (11,15 %) en comparación con los autores (13,72 % y 9,31 %, respectivamente).

Para los autores, la suma de las emociones negativas (MIEDO, IRA, TRISTEZA y AVERSIÓN) es de 49,76 % y para las periodistas mujeres es de 46,48 %, es decir, los primeros están más de 3 puntos por encima, si bien con una predominancia destacable de la emoción CONFIANZA para los dos grupos.

Exponemos a continuación los mismos datos, pero ahora obtenidos para los corpus B1 y B2, que contienen exclusivamente las secuencias humorísticas:

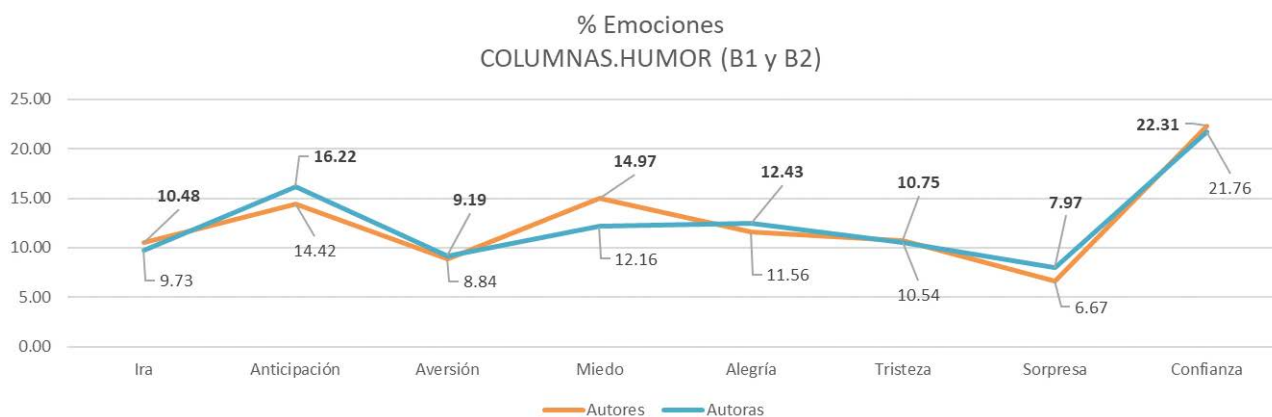


Figura 8. Emociones en los corpus B1 y B2 en porcentajes

Según los datos anteriores (Figura 8), la emoción más predominante tanto para los autores como para las autoras en el corpus de secuencias humorísticas vuelve a ser la CONFIANZA, con un valor del 22,31 % y 21,76 %, respectivamente, si bien aparece más predominantemente ahora en los autores. En cuanto a las emociones en las siguientes posiciones encontramos que:

- Para los autores, la segunda emoción más prevalente es el MIEDO (14,97 %) seguida de la ANTICIPACIÓN (14,42 %), la ALEGRÍA (11,56 %), la TRISTEZA (10,75 %), la IRA (10,48 %), la AVERSIÓN (8,84 %), y finalmente la SORPRESA (6,67 %).
- Para las autoras, después de la CONFIANZA la emoción más prevalente es la ANTICIPACIÓN (16,22 %), seguida de la ALEGRÍA (12,43 %), el MIEDO (12,16 %), la TRISTEZA (10,54 %), la IRA (9,73 %), la AVERSIÓN (9,19 %) y, finalmente, la SORPRESA (7,97 %).

Observamos algunas diferencias destacables entre autores y autoras. En los autores, después de la CONFIANZA (mayor en este corpus en comparación con las autoras) la segunda emoción más prevalente es el MIEDO, claramente negativa, mientras que en las autoras es la ANTICIPACIÓN, que puede ser etiquetada de manera positiva o negativa. La TRISTEZA, si bien es ligeramente mayor en los autores, la diferencia no es significativa y, sin embargo, la IRA sí es más prevalente en los autores tanto en este corpus, como en el anterior. Todas las emociones negativas tienen un porcentaje superior en los autores, salvo la de AVERSIÓN cuyo porcentaje es mayor en ellas (0,35 puntos por encima); la suma de los porcentajes de MIEDO, TRISTEZA, IRA y AVERSIÓN para los autores es 45,24 % frente al 41,62 % en las autoras. Se puede observar que, aunque no hay grandes diferencias, las autoras parecen tener una mayor proporción de emociones positivas, como la ALEGRÍA y la CONFIANZA, mientras que los autores tienden a mostrar más MIEDO, IRA y TRISTEZA.

Compararemos a continuación los corpus para detectar si en la expresión del humor destaca alguna característica emocional. La siguiente tabla contiene los porcentajes de las emociones del corpus con los artículos de opinión de los autores (A1) y del corpus que incluye exclusivamente sus secuencias humorísticas (B1); se destacan en rojo los valores más altos:

	Ira	Anticipación	Aversión	Miedo	Alegría	Tristeza	Sorpresa	Confianza
A1	12,56	13,72	8,93	16,22	9,31	12,05	6,02	21,18
B1	10,48	14,42	8,84	14,97	11,56	10,75	6,67	22,31

Tabla 4. Emociones de los corpus A1 y A2 en porcentajes

La Tabla 4 muestra algunas diferencias notables en las emociones utilizadas en los dos corpus estudiados de los autores. En el corpus de artículos de opinión completos (A1), la emoción más prevalente después de la CONFIANZA es el MIEDO (16,22 %), mientras que en el corpus de secuencias humorísticas (B1), la segunda emoción más prevalente es la ANTICIPACIÓN (14,42 %). Además, en el corpus de secuencias humorísticas (B1), la ALEGRÍA tiene un valor más alto (11,56 %) en comparación con el corpus de artículos de opinión completos (9,31 %), mientras que la

TRISTEZA tiene un valor más bajo (10,75 % frente a 12,05 %, respectivamente). Por lo tanto, parece que el humor, expresado en todas las secuencias, sí que es detectado a partir de las emociones que se expresan y, por lo tanto, deben considerarse en los análisis de emociones.

	Ira	Anticipación	Aversión	Miedo	Alegría	Tristeza	Sorpresa	Confianza
A2	11,62	14,54	8,73	14,33	11,15	11,80	6,55	21,29
B2	9,73	16,22	9,19	12,16	12,43	10,54	7,97	21,76

Tabla 5. Emociones de los corpus A2 y B2 en porcentajes

En la Tabla 5 podemos observar que para ambos corpus la emoción predominante sigue siendo la CONFIANZA, si bien es ligeramente mayor en el corpus de secuencias humorísticas (B2). También en este son más elevadas las emociones de ANTICIPACIÓN, ALEGRÍA y SORPRESA (más de un punto por encima en las tres emociones). Las emociones negativas se presentan más elevadas en el corpus A1 (como sería lo previsible por el tipo de texto que contiene), donde vemos que el MIEDO, la TRISTEZA y la IRA son superiores. En esta última tabla, el único sentimiento negativo más elevado en el corpus B2 es la AVERSIÓN, si bien no llega a medio punto (0,46).

6. Conclusiones

Este estudio se basaba en tres hipótesis, dos confirmadas y otra solo en parte. En primer lugar, pronosticamos que los corpus de secuencias humorísticas (A2 y B2) iban a contener sentimientos más intensos y predominantemente de carácter positivo, en comparación con los corpus con los artículos de opinión completos (A1 y B1). Aquí vimos que los valores promedio del TSS mostraron una mayor carga emocional en los textos humorísticos; sin embargo, no ocurrió lo mismo con el TSI, que salía más elevado para los corpus A1 y A2. Este resultado podría explicarse por el uso de palabras emocionales que tienen una polaridad más fuerte al tratar temas que despiertan emociones más intensas, como corresponde al artículo de opinión. En las secuencias humorísticas, por el contrario, los autores pueden utilizar un lenguaje más ligero y entretenido, lo que disminuye la puntuación del TSI. Los datos sugieren que las secuencias humorísticas tienen un mayor contenido emocional en general, pero una menor polarización en cuanto a la intensidad del sentimiento expresado. Si comparamos por género, la puntuación global de sentimiento y de intensidad al contrastar los corpus A1 y A2 era siempre ligeramente superior en las autoras; sin embargo, en las secuencias humorísticas, los valores de polarización e intensidad eran superiores en los autores. En el caso del corpus de artículos de opinión, esta mayor polaridad e intensidad podría deberse a que las autoras tienden a ser más expresivas y detalladas en sus narraciones, lo que resulta en una mayor presencia de palabras emocionales en general. Además, las autoras podrían estar más interesadas en abordar temas que involucren emociones o estar más dispuestas a expresar sus sentimientos en sus escritos. En el caso de las secuencias humorísticas, podría deberse a que los autores masculinos tuvieran una tendencia a usar un humor más mordaz o sarcástico, lo que puede resultar en una mayor polarización emocional, pero es una suposición, habría que contrastarlo con más datos. Por otro lado, las autoras podrían utilizar un humor más sutil y moderado, lo que resultaría en valores de polarización e intensidad más bajos.

En segundo lugar, la hipótesis pronosticaba que existía una mayor presencia de emociones negativas, como la TRISTEZA y la IRA en los textos completos a causa del tipo de texto que se trata, el artículo de opinión, a diferencia de lo que ocurrirá en el corpus de secuencias humorísticas, en donde el humor se utiliza para suavizar el tono general del artículo y vimos, en lo que respecta a emociones que, al comparar los corpus, en todos destaca la de CONFIANZA, que se caracteriza por ser predominantemente positiva, pues implica sentimientos de seguridad. Esta predominancia surge de los rasgos y patrones lingüísticos de los textos con una mayor frecuencia de palabras relacionadas con esta emoción, y así lo detectó EmoLex. Esto podría deberse también a la naturaleza de los temas y a las perspectivas presentadas por los autores. En cualquier caso, se necesitaría realizar un análisis más detallado del contenido específico y las características lingüísticas y de contexto para justificar con rigurosidad la prevalencia de esta emoción. El género de los autores o la orientación humorística o no de los corpus no identifica una correlación con respecto a esta emoción de CONFIANZA, dado que siempre ha sido la predominante. Por otra parte, con respecto a los autores masculinos, la suma de las emociones negativas, tanto en el corpus A1 como en el B1 siempre han sido superiores a la de las autoras, lo que coincide con lo apuntado por otros estudios que hemos mencionado.

La tercera hipótesis predecía que los autores y las autoras expresan emociones de manera similar, pero con una posible mayor presencia de emociones positivas en los textos de las autoras. Al comparar los corpus A1 y B1, apreciamos que no hay diferencias significativas y en ambos se evoca un sentimiento negativo, con una intensidad extremadamente alta en los dos casos, si bien ligeramente superior para las autoras; aquí, por tanto, la hipótesis no se cumplía. En el caso de los corpus con secuencias humorísticas, todas las emociones negativas tenían un porcentaje superior en los autores, salvo la de AVERSIÓN, que era mayor para las autoras. Se pudo observar que, aunque las diferencias no eran significativas, las autoras manifestaban una mayor proporción de emociones positivas, como la ALEGRÍA y la CONFIANZA, mientras que los autores tendieron a mostrar más MIEDO, IRA y TRISTEZA.

En resumen, aunque la presencia de polaridad y emociones negativas es alta en ambos géneros y corpus, las autoras parecen tener una tendencia a utilizar más emociones positivas en su escritura humorística que los autores.

En definitiva, este estudio nos ha permitido explorar cómo se manifiestan los sentimientos y las emociones en el lenguaje escrito de determinados periodistas en sus artículos de opinión y en sus secuencias humorísticas. Si bien nuestros resultados no cumplieron parte de una hipótesis, nos han brindado nuevas perspectivas sobre cómo los periodistas estudiados expresan y transmiten emociones a través de sus textos. A pesar de que el análisis de sentimiento y emociones en el lenguaje natural todavía es un campo en desarrollo, estos hallazgos pueden ser un primer paso para comprender mejor cómo las emociones influyen en la comunicación escrita de los artículos de opinión y en sus expresiones humorísticas, y cómo podemos utilizar este conocimiento para mejorar nuestra comprensión de la polaridad, la intensidad y las emociones. Como trabajo futuro, será necesario profundizar en estos resultados y complementarlos con otras técnicas de análisis de texto para seguir avanzando en nuestro entendimiento del complejo mundo de las emociones y del humor en el lenguaje escrito.

Agradecimientos

Este trabajo es resultado del proyecto PROMETEO/2021/079 “Etiquetaje pragmático para un observatorio de la identidad de mujeres y hombres a través del humor. La plataforma Observahumor.com” (Generalitat Valenciana).

Referencias bibliográficas

- Ahmadi-Azad, S. (2015). Gender differences in emotional content of EFL written narratives. *Journal of Language Teaching and Research*, 6(3), 619-626. <https://doi.org/10.17507/jltr.0603.19>
- Attardo, S. (1994). *Linguistic Theories of Humor*. De Gruyter Mouton.
- Bamman, D., Eisenstein, J. y Schnoebelen, T. (2012). Gender in Twitter: Styles, stances, and social network. <https://arxiv.org/abs/1210.4567v1>
- Banchs, R. E. (2017). On the Construction of more Human-like Chatbots: Affect and Emotion Analysis of Movie Dialogue Data. *Proceedings of APSIPA Annual Summit and Conference*, 1364-1367.
- Barrett, L. F., Robin, L., Pietromonaco, P. R., y Eyssell, K. M. (1998). Are Women the “More Emotional” Sex? Evidence from Emotional Experiences in Social Context. *Cognition and Emotion*, 12(4), 555-578. <https://doi.org/10.1080/026999398379565>
- Brody, L. R., y Hall, J. A. (2000). Gender, emotion, and expression. En M. Lewis y J. M. Haviland-Jones (Eds.), *Handbook of emotions* (pp. 338-349). Guilford Press.
- Cameron, D. (2005) Language, Gender, and Sexuality: Current Issues and New Directions, *Applied Linguistics*, 26(4), 482-502. <https://doi.org/10.1093/applin/ami027>
- Chaplin, T. M., y Aldao, A. (2013). Gender differences in emotion expression in children: A meta-analytic review. *Psychological Bulletin*, 139(4), 735-765. <https://doi.org/10.1037/a0030737>
- Cowie, R. (2023). Computational research and the case for taking humor seriously. *Humor*. <https://doi.org/10.1515/humor-2023-0021>
- Fischer, A. H. (1993). Sex Differences in Emotionality: Fact or Stereotype? *Feminism & Psychology*, 3(3), 303-318. <https://doi.org/10.1177/0959353593033002>
- Fischer, A. H., y Manstead, A. S. (2000). The relation between gender and emotions in different cultures. *Gender and emotion: Social psychological perspectives*, 1, 71-94.
- Gilligan, C. (2009). *In a Different Voice: Psychological Theory and Women's Development*. Harvard University Press.
- Goldshmidt, O. T., y Weller, L. (2000). Talking Emotions: Gender Differences in a Variety of Conversational Contexts. *Symbolic Interaction*, 23(2), 117-134. <https://doi.org/10.1525/si.2000.23.2.117>
- Kalloniatis, A. y Adamidis, P (2023). Computational Humor Recognition: A Systematic Literature Review, 07 February 2023, PREPRINT (Version 1) disponible en *Research Square*. <https://doi.org/10.21203/rs.3.rs-2552754/v1>
- Kilgariff A, Rychlý P, Smrž P, Tugwell D (2004). The sketch engine. En G. Williams y S. Vessier (Eds.). *Proceedings of the 11th EURALEX international congress*, (pp 105-116). Université de Bretagne-Sud, Faculté des Lettres et des Sciences Humaines.
- López-Rodríguez, C. I. (2022). Emotion at the end of life: Semantic annotation and key domains in a pilot study audiovisual corpus. *Lingua* 277. <https://doi.org/10.1016/j.lingua.2022.103401>
- Mahajan, R., y Zaveri, M.A. (2020). Humor identification using affect-based content in target text. *J. Intell. Fuzzy Syst.*, 39, 697-708.
- Martínez Egido, J. J. (2014). El humor en el artículo de opinión. *Feminismo/s*, 24. <https://doi.org/10.14198/fem.2014.24.06>
- Mihalcea, R., y Strapparava, C. (2006). Technologies that make you smile: Adding humor to text-based applications. *IEEE Intelligent Systems*, 21(5), 33-39. <https://doi.org/10.1109/MIS.2006.104>
- Mohammad, S. M., y Turney, P. D. (2013). Crowdsourcing a word-emotion association lexicon. *Computational intelligence*, 29(3), 436-465.
- Mohammad, S. M. (2018). Word Affect Intensities. *Proceedings of the 11th edition of the Language Resources and Evaluation Conference*, mayo 2018, Miyazaki, Japón.
- Mohammad, S. M. (2020). Practical and Ethical Considerations in the Effective use of Emotion and Sentiment Lexicons. <http://arxiv.org/abs/2011.03492>
- Moreno-Ortiz, A. (2021). Lingmotif 2.0 [software]. Universidad de Málaga. <https://tlt.uma.es>
- Moreno-Ortiz, A. (2017). Lingmotif: A User-focused Sentiment Analysis Tool. *Procesamiento del Lenguaje Natural*, 58(0), 133-140.

- Nandwani, P. y Verma, R. (2021) A review on sentiment analysis and emotion detection from text. *Social Network Analysis and Mining*, 11, 81. <https://doi.org/10.1007/s13278-021-00776-6>
- Ortega-Gilabert, J. A., y Timofeeva-Timofeev, L. (2023). Claves del etiquetaje pragmático en el corpus CHILDHUM. *Círculo de Lingüística Aplicada a la Comunicación* 96, 59-71. <https://dx.doi.org/10.5209/calcul.91591>
- Orts, M. Á. (2021). Emotion to forecast a recession: A bilingual lexical and sentiment analysis of the un and IMF world economy reports for 2019. *Iberica*, 40, 217-244.
- Plutchik, R. (2001) The nature of emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. *American scientist*, 894, 344-350.
- Plutchik, R. y Kellerman, H. (Eds.) (2013). *Theories of emotion* (Vol. 1). Academic Press.
- Taboada, M. (2016). Sentiment Analysis: An Overview from Linguistics. *Annual Review of Linguistics*, 2, 325-347. <https://doi.org/10.1146/annurev-linguistics-011415-040518>
- Raskin, V. (1985). *Semantic mechanisms of humor*. Reidel
- Reyes, A. (2013). Linguistic-based Patterns for Figurative Language Processing: The Case of Humor Recognition and Irony Detection. *Procesamiento del Lenguaje Natural*, 50, 107-109.
- Reyes, A. y Rosso, P. (2009). Linking Humour to Blogs Analysis: Affective Traits in Posts. *Proceedings of the 1st Workshop on Opinion Mining and Sentiment Analysis (WOMSA), CAEPIA-TTIA Conference*, 100-109. <http://users.dsic.upv.es/grupos/nle/downloads.html>
- Reyes, A., Buscaldi, D. y Rosso, P. (2009a). The Impact of Semantic and Morphosyntactic Ambiguity on Automatic Humour Recognition. *NLDB*, 130-141.
- Reyes, A., Rosso, P. y Buscaldi, D. (2009b). Affect-based features for humour recognition. *Proceedings of the 7th International Conference on Natural Language Processing ICON-09*.
- Ruiz Gurillo, L. (2019). En modo humorístico: habilidades metapragmáticas en discursos de humor. En Antonio Briz Gómez et al. (coords.), *Estudios lingüísticos en homenaje a Emilio Ridruejo*, 2 vols., (pp. 1231-1242). Universitat de València.
- Taboada, M. (2016). Sentiment analysis: an overview from linguistics. *Annu Rev Linguist*, 2, 325-347. <https://doi.org/10.1146/annurev-linguistics-011415-040518>
- van den Beukel, S. y Lora Aroyo (2018). Homonym Detection for Humor Recognition in Short Text. *Proceedings of the 9th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, (pp. 286-291). Association for Computational Linguistics.
- Vainik, E. (2006). Intracultural variation of semantic and episodic emotion knowledge in Estonian. *TRAMES*, 10(60/55), 2, 169-189.
- Vargas-Sierra, C. y Orts, M.A. (2023). Sentiment and Emotion in Financial Journalism: A Corpus-based, Cross-linguistic Analysis of the Effects of COVID. *Humanit Soc Sci Commun* 10, 219. <https://doi.org/10.1057/s41599-023-01725-8>
- Zhang, D., Song, W., Liu, X., Liu, L., y Zhao, X. (2018). Research on Humor Recognition. *IEEE 9th International Conference on Software Engineering and Service Science (ICSESS)*, 152-155.