

**CREACIÓN Y EVALUACIÓN DE UN DICCIONARIO
MARCADO CON EMOCIONES Y PONDERADO
PARA EL ESPAÑOL**

**CREATION AND EVALUATION
OF TAGGED WITH EMOTIONS AND WEIGHTED DICTIONARY
FOR SPANISH**

Ismael Díaz Rangel, Grigori Sidorov, Sergio Suárez-Guerra

Instituto Politécnico Nacional, CIC-IPN, México

Resumen

Este artículo presenta un método para la creación de diccionarios marcados con un valor específico (por ejemplo, las emociones, la polaridad) para su uso en varias tareas de procesamiento de lenguaje natural realizadas por computadoras. En el diccionario creado las palabras seleccionadas se han etiquetado con seis emociones básicas. Para eso, las palabras primero fueron analizadas (anotadas) manualmente por múltiples evaluadores, y ponderadas automáticamente a base de éstas. El método se aplicó para el idioma español. Las palabras que conforman al diccionario fueron etiquetadas con las categorías emocionales básicas: alegría, enojo, miedo, tristeza, sorpresa y repulsión. A diferencia de otros diccionarios para computadoras, el diccionario propuesto contiene ponderaciones —porcentajes de probabilidad de ser usadas con un sentido emocional. Cada palabra fue valorada por múltiples evaluadores, y posteriormente se realizó un análisis de concordancia con el método de kappa ponderado, adaptándolo para evaluadores múltiples. Con los resultados obtenidos, se propuso una medida que estima que tan frecuente es el uso afectivo de una palabra: factor de probabilidad de uso afectivo (FPA), el cual sirve para dotar a las palabras potencialmente emocionales con un factor de peso. El FPA puede ser incluido como información en sistemas automáticos, por ejemplo, para detección de sentimientos en texto. El FPA se refiere a la tendencia del uso de cada palabra, no es una característica absoluta. Así es útil para los sistemas automáticos.

Palabras clave: diccionario marcado con emociones, factor de probabilidad de uso afectivo, concordancia entre evaluadores, análisis de sentimientos, método de kappa ponderado.

Abstract

This paper presents a method for creation of dictionaries marked with specific values (for example, emotions, polarity) for use in various tasks of automatic natural language processing. In the created dictionary, the selected words are tagged with six basic emotions. For this, they are first analyzed (annotated) manually by multiple annotators and automatically weighted on the basis of these evaluations. The method was applied to the Spanish language. The paradigm chosen for tagging of the words that form the dictionary corresponds to basic emotional categories: joy, anger, fear, sadness, surprise and disgust. Unlike other dictionaries, our dictionary contains weights that correspond to percentages of probability of being used with the sense related to emotion. Each word was evaluated by multiple annotators, and subsequently the agreement between them was analyzed with the method of weighted kappa adapted for multiple entries. On the basis of these results, we propose a new measure that estimates the probability of the affective use: probability factor of affective use (PFA), which serves to provide the potentially emotional words with the weight. PFA can be used as data in automatic systems for emotional analysis of texts. PFA refers to the tendency of the use of each word, which is useful for automatic systems.

Key words: emotional dictionary, probability factor of affective use, agreement between annotators, sentiment analysis, method of weighted kappa.

I. Introducción

La detección automática de sentimientos en texto cada día cobra un mayor interés en diversos sectores; en parte se debe a la inmensa cantidad de información disponible en la red como son análisis o evaluaciones de: productos de consumo, programas informáticos, deportes, películas, etcétera; todo ello accesible en sitios especializados como revistas electrónicas, blogs y foros; son también una importante fuente de información los sitios de ventas, por ejemplo, *Amazon*. A su vez, es de creciente interés la identificación de contenido emocional presente en sitios de noticias y redes sociales como *Facebook* y *Twitter*.

La información a que se hace referencia es una vasta y útil fuente de datos para su procesamiento automático por computadoras. Se puede aplicar, por ejemplo, para mercadología, asesoría de imagen y política, psicología, seguridad —a nivel desde personal

hasta nacional—, sector lúdico, interacción hombre-máquina, síntesis de voz y cualquier área interesada en la extracción y conocimiento de opiniones, estados de ánimo, observaciones y evaluaciones.

Las principales áreas que se encargan de clasificar automáticamente palabras, textos o documentos de acuerdo a la opinión, emoción o sentimiento que expresan, son conocidas como Minería de opiniones (*Opinion mining*) y Análisis de sentimientos (*Sentiment analysis*).

“Existen diversos problemas inherentes en la detección de sentimientos, como muestra basta mencionar que casi cualquiera palabra, potencialmente puede tener contenido afectivo, incluso aquellas que pueden parecer neutrales” (Strapparava y Mihalcea, 2007); algunas palabras en cualquier contexto o casi en cualquier contexto denotan o sugieren una emoción, un ejemplo puede ser la palabra “regocijo”, en cuyo caso todos sus sentidos son relacionados con una emoción; sin embargo, otras palabras dependen del contexto, como, digamos, la palabra “depresión”: algunos sentidos sí son relacionados con las emociones, mientras que otros no. “Hay casos más complejos, como la palabra “fantasma”, donde el potencial afectivo es parte de la imaginación colectiva” (Strapparava y Mihalcea, 2007).

Existen diversos enfoques con los que se puede realizar la clasificación o identificación de emociones en un texto dado, los más relevantes según un estudio realizado (Cowie y Cornelius, 2003) son: modelos circunflejos (Russell, 1980), descripciones basadas en evaluación (Scherer, 1984), dimensiones emocionales (Wundt, 1896) y categorías emocionales (Smith, 1989).

Es más común encontrar trabajos relacionados con los dos últimos enfoques: las **dimensiones emocionales** consideran el nivel presente sobre tres aspectos: valencia (o polaridad), activación y control. La polaridad indica si existe un estado emocional positivo o negativo, donde la felicidad se considera positiva y la tristeza negativa. La activación indica una situación activa o pasiva del estado emocional, donde el extremo activo puede ser la excitación y el pasivo la calma. El control indica por un lado la dominación y al otro extremo la sumisión.

Con respecto a las **categorías emocionales**, se refiere a ubicar la unidad de estudio (palabras, oraciones, párrafos, textos, etcétera) dentro de la clasificación propuesta en alguna teoría de emociones, como las observadas en la tabla 1 (Ortony y Turner, 1990).

Tabla 1. Algunas teorías de clasificación de emociones básicas.

INVESTIGADOR	EMOCIONES BÁSICAS	BASES
Arnold	Ira, aversión, coraje, abatimiento, deseo, desesperación, miedo, odio, esperanza, amor, tristeza	Relación con las tendencias de actuar
Ekman, Friesen y Ellsworth	Ira, repulsión, miedo, alegría, tristeza, sorpresa	Expresiones fisiológicas faciales universales
Gray	Rabia, terror, ansiedad, alegría	Involuntarias
McDougall	Ira, asco, alegría, miedo, sumisión, oferta de emoción, asombro	Relación a los instintos

En Minería de opiniones y Análisis de sentimientos —como en muchas otras áreas—, para lograr los objetivos buscados son necesarios diversos recursos lingüísticos; los recursos léxicos son indispensables y existen varios disponibles para el idioma inglés, como el *SentiWordNet* (Esuli y Sebastiani, 2006), el *General Inquirer* (Stone y otros, 1996) y el ANEW (*Affective Norms for English Words*) (Redondo y otros, 2007); en general los recursos para el idioma español son escasos, y en cuestiones puntuales inexistentes. Como ejemplos de recursos para determinar la polaridad de opiniones en idioma español están: el *Sentitex* (Moreno y otros, 2010a), que entre otras cosas se ha utilizado para analizar la opinión de usuarios acerca de hoteles (Moreno y otros, 2010b) y para realizar el análisis de emociones en redes sociales (Moreno y Pérez, 2013); también existe el recurso *EmotiBlog* (Boldrini y otros, 2009) que incluye los idiomas español, inglés e italiano. Sin embargo, esos recursos para el español no están marcados con emociones, y solamente tienen valores positivo o negativo para el análisis de polaridad.

Este trabajo está centrado en la metodología para la creación de diccionarios con palabras emocionales que fue aplicada para el idioma español; se diferencia en varios aspectos de la adaptación al español del ANEW; primero por el hecho de que cada palabra proporciona información acerca de su categoría emocional, tomando como referencia las seis emociones básicas propuestas por Ekman (1972), mientras que ANEW proporciona información con un enfoque de dimensiones emocionales; segundo, las palabras incluidas indican la frecuencia con que se usan con un sentido emocional, considerando diferentes contextos de aplicación; tercero, se aplicó una adaptación del método de concordancia ponderada en las evaluaciones realizadas, mientras que ANEW no lo hace.

Dado que la mayor cantidad de recursos están hechos para el idioma inglés, es importante desarrollarlos para otros idiomas, específicamente para el español, donde tan solo en América Latina existen más de 235 millones de usuarios de Internet (internautas¹).

El trabajo tiene la estructura siguiente. En el apartado 2, se explica el criterio de selección de palabras y su colocación en una categoría emocional específica. En el apartado 3, se indica cómo con la ayuda de múltiples evaluadores, se realizaron estimaciones acerca del contenido emocional de las palabras seleccionadas, siguiendo una serie de criterios de valoración y reglas de etiquetado. En el apartado 4, se realiza un análisis de las evaluaciones utilizando el método de kappa ponderado, haciendo adecuaciones para su aplicación sobre múltiples evaluadores. En la sección 5, se escogen los datos de los evaluadores más afines usando la información obtenida en el análisis de concordancia entre ellos, para calcular el factor de probabilidad de uso afectivo (FPA) asociado a cada palabra. Al final, se presentan conclusiones.

2. Selección de palabras del diccionario

Nuestra propuesta consiste en creación de un diccionario etiquetado con emociones básicas: alegría, enojo, miedo, repulsión, sorpresa y tristeza. En primera instancia se tomó el recurso *WordNet-Affect* (Strapparava y Valitutti, 2004) para el inglés, el cual está conformado por *synsets*, en este caso son grupos de palabras sinonímicas que comparten valores en su etiqueta la cual tiene tres aspectos: positivo, negativo y objetivo, cuya suma es 1. “Este recurso fue creado usando varias técnicas lingüísticas y clasificadores estadísticos, fue construido semiautomáticamente, y no todos los resultados fueron revisados manualmente, por lo que son de esperarse algunas clasificaciones incorrectas” (Martín y otros, 2010). Por otro lado, las palabras de este recurso también están separadas por categorías emocionales bajo el enfoque de Ekman.

Para la creación del diccionario propuesto se tomaron todas las palabras incluidas en cada categoría del recurso *WordNet-Affect*, con los resultados mostrados en la tabla 2.

Tabla 2. Palabras por categoría obtenidas del *WordNet-Affect*.

CATEGORÍA	PALABRAS (INGLÉS)
<i>Alegría</i>	399

¹ <http://www.internetworldstats.com/stats.htm>

<i>Repulsión</i>	252
<i>Enojo</i>	137
<i>Miedo</i>	53
<i>Sorpresa</i>	67
<i>Tristeza</i>	202
Total	1,110

Dado que *WordNet-Affect* es un recurso en inglés, el siguiente paso fue traducir automáticamente cada una de las palabras, preservando su etiqueta de categoría emocional; para ello se utilizaron tres herramientas: traductor de *Google*, el *English-Spanish Interpreter Pro* y el traductor *Babylon*. La tabla 3 es un ejemplo de los resultados obtenidos.

Tabla 3. Segmento con traducción de palabras.

PALABRA	GOOGLE	INTERPRETER	BABYLON
<i>adorably</i>	<i>adorablemente</i>	<i>adorablemente</i>	<i>adorablemente,</i> <i>encantadoramente</i>
<i>adoration</i>	<i>adoración</i>	<i>adoración</i>	<i>adoración,</i> <i>pleitesía,</i> <i>veneración</i>

Lo siguiente consistió en integrar los resultados de los tres traductores automáticos y eliminar los duplicados; la cantidad de términos resultantes por categoría emocional se pueden observar en la tabla 4.

Tabla 4. Cantidad de palabras por categoría después de la traducción.

CATEGORÍA	PALABRAS (ESPAÑOL)
<i>Alegría</i>	1,177
<i>Repulsión</i>	206
<i>Enojo</i>	959
<i>Miedo</i>	431
<i>Sorpresa</i>	267
<i>Tristeza</i>	551
Total	3,591

Sin embargo, algunos resultados de la traducción deben ser descartados; por ejemplo, en la categoría “Tristeza” se encuentra la palabra inglesa “blue”, el traductor entregó “tristeza”, pero también “azul”; en español este término no tiene ninguna relación con la tristeza, por lo que debe ser suprimido. Para garantizar la calidad de las palabras se realizó una revisión

manual de cada una de las 3,591 palabras. Todas las acepciones de cada palabra —presentes en el diccionario de María Moliner (1996)— fueron verificadas manualmente para ver si existe alguna relación con las emociones básicas: el criterio fue mantener aquellas palabras que en alguna de sus acepciones el significado indicara relación con la emoción de la categoría asociada; ante la menor duda en el análisis, la palabra se conservó. Esto representó un laborioso trabajo manual, con la finalidad de obtener un recurso léxico de alta calidad, lo cual puede derivar en un mejor desempeño al utilizarlo en un sistema automático de reconocimiento de emociones; existen trabajos donde una idea similar se ha llevado a cabo para otros conjuntos de palabras, como en lo expuesto por (Moreno y otros, 2011).

Después del minucioso análisis y depuración de términos, se agregaron las palabras clasificadas según el criterio de Ekman mostradas en Greenberg (2000), el cual muestra las emociones primarias y las emociones secundarias más cercanas. La cantidad total resultante de palabras fue de 2,036 como se observa en la Tabla 5.

Tabla 5. Tamaño del diccionario después del depurado.

CATEGORÍA	PALABRAS (ESPAÑOL)
<i>Alegría</i>	668
<i>Repulsión</i>	209
<i>Enojo</i>	382
<i>Miedo</i>	211
<i>Sorpresa</i>	175
<i>Tristeza</i>	391
Total	2,036

3. Apreciación de evaluadores sobre emociones en palabras del diccionario

Una vez definida la lista de palabras del diccionario, se procedió a evaluar si las palabras propuestas guardaban relación con la emoción con que fueron asociadas —etiquetadas en la etapa de selección—. Para ello se pidió ayuda a evaluadores voluntarios. Hay varias maneras de pedir que se realicen evaluaciones, lo cual depende de la información que se desee obtener. A los evaluadores, además del diccionario, se les proporcionó un instructivo de evaluación, el cual se puede resumir así:

“Lo que se busca es cuantificar que tan probable es que el sentido con el que se usa la palabra que se evalúa, denote relación con la emoción a la cual se asocia. Nótese que **no se evalúa si el significado de la palabra está relacionado con alguna emoción o la intensidad de la emoción, sino más bien, es la estimación de que tan frecuente se usa esta palabra para denotar algo relacionado con dicha emoción, imaginando sus posibles contextos de uso y su frecuencia.**” Esto se hace marcando una de las opciones que tiene cada palabra: “Nula”, “Baja”, “Media” y “Alta”.

“Nula” indica que bajo ningún contexto el significado de la palabra denota relación con la emoción a la cual se asocia; “Baja” indica que en algunos contextos el significado de la palabra sí denota algo relacionado con la emoción, pero es poco frecuente usarla con ese propósito; “Media” para indicar que con igual frecuencia la palabra se usa o no para denotar algo relacionado con la emoción a la cual se asocia; y por último, “Alta”, se usa para indicar que casi —o siempre— la palabra denota algo relacionado con la emoción asociada.

Dado que algunos términos pueden ser poco usuales, es posible que el evaluador no recuerde o no conozca los posibles contextos en los cuales el término se puede usar, se procedió a colocar manualmente las definiciones —muchas de ellas con oraciones ejemplo— a casi la mitad de los términos incluidos, para ello se utilizó la información del diccionario de uso del español de María Moliner; para acceder a ellas, el evaluador solo debe pasar el cursor sobre el término y automáticamente se despliegan sus definiciones, tabla 6.

Tabla 6. Segmento de lista de palabras etiquetadas con la emoción “Miedo”.

1	PALABRA	NULA	BAJA	MEDIA	ALTA		
2	abominable						
3	accidente						
4	acobardar						
5	acomplejado						
6	acoquinamiento						
7	acoquinar	tr. inf. Hacer que alguien tenga miedo o esté paralizado o sometido por el miedo: 'No consiste ser buen maestro en tener acoquinados a los niños'. 1 *Intimidar. 5 prnl. inf. Amilanarse, acobardarse.					
8	agüero						
9	ahuyentar						
10	alarma						

Es de mencionar que cada evaluador trabajó íntegramente sobre el conjunto de palabras de una categoría emocional. La tabla 7 muestra cuantos evaluadores participaron en la valoración de los términos sobre cada categoría emocional.

Tabla 7. Evaluaciones realizadas en las palabras de cada categoría emocional.

CATEGORÍA	PALABRAS	EVALUADORES
<i>Alegría</i>	668	19
<i>Repulsión</i>	209	17
<i>Enojo</i>	382	17
<i>Miedo</i>	211	19
<i>Sorpresa</i>	175	19
<i>Tristeza</i>	391	17
Total	2,036	108

Se muestra un ejemplo de los datos antes de que se realizara una depuración de ellos. La tabla 8 contiene un fragmento de las estimaciones de los 19 evaluadores de la categoría “Alegría”. En ella se observa la palabra estimada y la cantidad de evaluadores que eligieron asociarla a una de las categorías propuestas; entre paréntesis se indica el porcentaje correspondiente. A simple vista se puede intuir el grado de concordancia/discordancia de las evaluaciones, a mayor porcentaje en una sola categoría mayor concordancia; aunque esto es sin considerar los efectos del azar, por lo que se aplicó un método que sí considera dichos efectos, lo cual se comenta más adelante.

Tabla 8. Estimaciones de evaluadores.

#	Palabra	Nula	Baja	Media	Alta
1	<i>abundancia</i>	1 (5.3%)	2 (10.5%)	8 (42.1%)	8 (42.1%)
2	<i>acabalar</i>	7 (36.8%)	1 (5.3%)	9 (47.4%)	2 (10.5%)
3	<i>acallar</i>	9 (47.4%)	8 (42.1%)	1 (5.3%)	1 (5.3%)
4	<i>acatar</i>	9 (47.4%)	5 (26.3%)	4 (21.1%)	1 (5.3%)
5	<i>acción</i>	6 (31.6%)	7 (36.8%)	4 (21.1%)	2 (10.5%)
6	<i>aceptable</i>	4 (21.1%)	3 (15.8%)	11 (57.9%)	1 (5.3%)
7	<i>aceptación</i>	3 (15.8%)	2 (10.5%)	6 (31.6%)	8 (42.1%)
8	<i>acicate</i>	4 (21.1%)	6 (31.6%)	7 (36.8%)	2 (10.5%)
9	<i>aclamación</i>	3 (15.8%)	2 (10.5%)	4 (21.1%)	10 (52.6%)

4. Análisis de evaluaciones

En las evaluaciones donde se involucren los estados afectivos la subjetividad podrá estar presente, por lo que se espera que existan algunas diferencias en las estimaciones entre un evaluador y otro; sin embargo, lo que nos interesa es saber dónde hay mayores coincidencias, ya que con ello se sabe que tanto se puede o no relacionar una palabra con un contexto afectivo.

Por otro lado, las discrepancias pueden estar dadas por la displicencia al realizar las evaluaciones, incluso es probable que algún evaluador realice al marcaje —integral o parcialmente— de manera aleatoria. Por lo que es necesario un mecanismo con el cual se pueda analizar la concordancia entre las evaluaciones realizadas, y descartar el trabajo de los evaluadores que presenten una notoria discordancia con respecto a los demás evaluadores.

En términos generales, la concordancia es el grado con que dos o más observadores, métodos, técnicas u observaciones están de acuerdo sobre el mismo fenómeno observado (Cortés, 2008). Por lo tanto, la concordancia no evalúa la validez o la certeza sobre observaciones con relación a un estándar de referencia dado, sino cuán afines están entre sí las observaciones sobre el mismo fenómeno.

El índice de acuerdo observado simplemente expresa en qué proporción hubo coincidencia en la clasificación entre los observadores, en relación al total de elementos examinados. Este se obtiene sumando la cantidad de objetos de estudio, donde hubo coincidencias en las evaluaciones, y se divide por el total de objetos analizados. El problema que plantea este índice básico es que una parte de ese acuerdo, puede deberse exclusivamente al azar.

El índice de kappa diseñado por Cohen (1960), ajusta el efecto del azar en la proporción de la concordancia observada. En esencia, el proceso de elaboración del índice es el siguiente: se calcula la diferencia entre la proporción de acuerdo observado y la proporción de acuerdo esperado por azar; si ésta es igual a cero, entonces el grado de acuerdo observado puede atribuirse enteramente al azar; si la diferencia es positiva, ello indica que el grado de acuerdo es mayor que el que cabría esperar si solo estuviera operando el azar.

Este índice es empleado exclusivamente para dos evaluadores y variables binarias —por ejemplo, positivo o negativo—, los resultados de las evaluaciones de los objetos de estudio

deben ser introducidos en una tabla de frecuencias de doble entrada o de contingencia (tabla 9).

Tabla 9. Tabla de contingencia.

		EVAL2		
		Positivo	Negativo	Total
EVAL1	Positivo	n_{11}	n_{12}	A_1
	Negativo	n_{21}	n_{22}	A_2
	Total	B_1	B_2	N

n_{11} es la cantidad de los objetos en los que el evaluador 1 y 2 han estimado como positivos².

n_{12} es la cantidad de objetos que el evaluador 1 calificó como positivos y el 2 como negativos.

n_{21} es la cantidad de objetos que el evaluador 1 calificó como negativos y el 2 como positivos.

n_{22} es la cantidad de los objetos que el evaluador 1 y 2 apreciaron como negativos.

A_1 y A_2 son la suma de los elementos de su fila ($A_1 = n_{11} + n_{12}$).

B_1 y B_2 son la suma de los elementos de su columna ($B_1 = n_{11} + n_{21}$).

N es la cantidad de objetos evaluados.

La estimación por el índice de kappa sigue la ecuación:

$$k = \frac{P_o - P_e}{1 - P_e} \quad (1)$$

donde:

P_o es la proporción de acuerdo observado y se calcula con:

$$P_o = \sum_{j=1}^2 \frac{n_{jj}}{N} \quad (2)$$

P_e es la proporción de acuerdo esperado al azar y se calcula con:

$P_e = \sum_{j=1}^2 \frac{A_j \cdot B_j}{N^2} \quad (3)$	
--	--

² Las sumatorias se realizan considerando la evaluación de un mismo objeto, esto aplica a las demás celdas.

es decir, en la forma completa:

$$P_e = \frac{((n_{11} + n_{12}) \cdot (n_{11} + n_{21})) + ((n_{21} + n_{22}) \cdot (n_{12} + n_{22}))}{N^2} \quad (4)$$

Una interpretación más popular propuesta por (Landis y Koch, 1977) para valorar el grado de acuerdo en función del índice de kappa se presenta en la tabla 10.

Tabla 10. Interpretación de índice de kappa.

KAPPA	GRADO DE ACUERDO
< 0.0	sin acuerdo
0.0 – 0.2	insignificante
0.2 – 0.4	bajo
0.4 – 0.6	moderado
0.6 – 0.8	bueno
0.8 – 1.0	muy bueno

En nuestro caso para la evaluación de las palabras que conforman al diccionario las variables no son binarias, son de tipo multinominal, es decir tienen más de dos opciones o categorías para su evaluación; además son variables ordinales; esto es, que tienen un orden establecido y no es intercambiable; por lo que, aunque no se dé una evaluación idéntica sobre un objeto, existirán valoraciones de concordancia diferentes, dependiendo de la posición de la categoría seleccionada en la evaluación. La tabla 11 contiene un ejemplo para la evaluación de múltiples categorías (Cat_x).

Tabla 11. Tabla de contingencia para kappa ponderado.

		EVAL 2				
		Cat 1	Cat 2	...	Cat m	Total
EVAL 1	Cat 1	n ₁₁	n ₁₂	...	n _{1m}	A ₁
	Cat 2	n ₂₁	n ₂₂	...	n _{2m}	A ₂

	Cat m	n _{m1}	n _{m2}	...	n _{mm}	A _m
	Total	B ₁	B ₂	...	B _m	N

Para este tipo de evaluaciones, un método de concordancia es el de kappa ponderado (Cohen, 1968), en el cual se establecen pesos para cuantificar la proporción de concordancia, en función de la distancia entre las *m* categorías elegidas en la evaluación del

objeto de estudio. Para ello, se asigna un peso (w_{ij}) a la celda de cada categoría (tabla 12), con un valor comprendido entre 0 y 1.

Tabla 12. Tabla de doble entrada para asignación de pesos

		EVAL 2			
		Cat 1	Cat 2	...	Cat m
EVAL 1	Cat 1	w_{11}	w_{12}	...	w_{1m}
	Cat 2	w_{21}	w_{22}	...	w_{2m}

	Cat m	w_{m1}	w_{m2}	...	w_{mm}

Los valores de los pesos dependen de la importancia que se quiera adjudicar a los desacuerdos. Inicialmente Cohen propuso una escala lineal (tabla 13a) de pesos dada por:

$$w_{ij} = 1 - \frac{|i - j|}{m - 1} \quad 1 \leq i, j \leq m \quad (5)$$

donde m es la cantidad de categorías.

Uno de los pesos más usados fue propuesto posteriormente en (Fleiss y Cohen, 1973), el cual propone una escala conocida como bicuadrática o cuadrática (tabla 13b):

$$w_{ij} = 1 - \frac{(i - j)^2}{(m - 1)^2} \quad (6)$$

Tabla 13. Pesos lineales y bicuadrados para 4 categorías.

a)

Peso Lineal		EVAL 2			
		Cat 1	Cat 2	Cat 3	Cat 4
EVAL 1	Cat 1	1	2/3	1/3	0
	Cat 2	2/3	1	2/3	1/3
	Cat 3	1/3	2/3	1	2/3
	Cat 4	0	1/3	2/3	1

b)

Peso Bicuadrado		EVAL 2			
		Cat 1	Cat 2	Cat 3	Cat 4
EVAL 1	Cat 1	1	8/9	5/9	0
	Cat 2	8/9	1	8/9	5/9
	Cat 3	5/9	8/9	1	8/9
	Cat 4	0	5/9	8/9	1

La expresión para la estimación del índice de kappa ponderada es:

$$k_w = \frac{P_{ow} - P_{ew}}{1 - P_{ew}} \quad (7)$$

Para calcular la concordancia observada (P_{ow}) se multiplica la tabla de evaluaciones por la tabla de pesos:

$$P_{ow} = \sum_{i=1}^k \sum_{j=1}^k w_{ij} \cdot \left(\frac{n_{ij}}{N}\right) \quad (8)$$

Para calcular la concordancia esperada (P_{ew}) se multiplican los límites marginales por la tabla de pesos:

$$P_{ew} = \sum_{i=1}^k \sum_{j=1}^k w_{ij} \cdot \left(\frac{A_i}{N}\right) \cdot \left(\frac{B_j}{N}\right) \quad (9)$$

El método de kappa ponderado es únicamente para dos evaluadores. Este trabajo tiene más evaluadores, sin embargo, es posible utilizar el kappa ponderado trabajando por parejas de evaluadores y al final hacer un promedio total con todos los valores obtenidos. Trabajar de esta manera también da la oportunidad de conocer el aporte —en cuanto a concordancia— de cada evaluador; para ello, solo se estima el promedio de los valores de kappa obtenidos entre dicho evaluador vs. todos los demás. Para saber cuántas parejas diferentes se pueden formar entre todos los evaluadores usamos:

$$\text{número de parejas} = \sum_{i=1}^{M-1} p + i \quad \forall M \geq 2 \quad (10)$$

donde M es el número de evaluadores.

La tabla 14 representa el contenido de las evaluaciones de kappa obtenidos de cada pareja ($k_w_eval\{i_j\}$); son un total de M por M celdas, aunque no todas son parejas distintas y la diagonal no cuenta. Para el caso de las palabras etiquetadas con la emoción “Alegría” participaron 19 evaluadores, por lo que se formaron 171 parejas realmente distintas de estimaciones de kappa.

Tabla 14. Tabla de estimación de kappa por pares.

Parejas 1er evaluador	Parejas 2º evaluador	...	Parejas M evaluador
$k_{w_eval} \{1_1\}$	$k_{w_eval} \{2_1\}$...	$k_{w_eval} \{M_1\}$
$k_{w_eval} \{1_2\}$	$k_{w_eval} \{2_2\}$...	$k_{w_eval} \{M_2\}$
...
$k_{w_eval} \{1_M\}$	$k_{w_eval} \{2_M\}$...	$k_{w_eval} \{M_M\}$

En este punto se procedió a obtener el índice de kappa promedio de cada evaluador, esto se realizó haciendo la sumatoria de todas las parejas de estimaciones de kappa que se formaron con cada evaluador (i), y dividiendo el resultado entre la cantidad de evaluadores menos uno.

$$k_{w_peval_i} = \frac{1}{M-1} \sum_{j=1}^M k_{w_eval}\{i_j\} \quad \forall i \neq j \quad (11)$$

Para conocer el grado de concordancia general o total de todas las evaluaciones sobre el diccionario se promedian todas las concordancias:

$$k_{w_total} = \frac{1}{M} \sum_{i=1}^M k_{w_peval_i} \quad (12)$$

Los resultados obtenidos para cada diccionario se muestran en la tabla 15, donde *lineal* y *bicadrado* se refiere al tipo de pesos usados en la estimación del índice de kappa.

Tabla 15. Resultados de estimación del kappa total por categoría emocional.

Categoría	k_{w_total}	
	Lineal	Bicadrado
<i>Alegría</i>	0.641	0.771
<i>Enojo</i>	0.649	0.781
<i>Miedo</i>	0.671	0.798
<i>Repulsión</i>	0.622	0.741
<i>Sorpresa</i>	0.641	0.76
<i>Tristeza</i>	0.647	0.779

5. Estimación de factor de probabilidad de uso afectivo (FPA)

La finalidad de realizar evaluaciones en el diccionario emocional propuesto, y aplicar a las evaluaciones un método de concordancia, fue la de conocer hasta qué punto—a pesar de la

subjetividad—, un término puede ser relacionado con una emoción en particular; el nivel de concordancia obtenido, nos da una referencia acerca de la validez de la hipótesis: que es posible cuantificar la frecuencia de uso afectivo de palabras sobre diferentes contextos.

Realizar las estimaciones de kappa por pares, nos da la oportunidad de conocer la contribución de cada evaluador de manera independiente, lo cual puede servir como mecanismo para la depuración de la evaluaciones, bajo el supuesto de que algunos evaluadores hayan trabajado de manera displicente. La tabla 16 muestra las estimaciones de kappa de cada evaluador, esto para la categoría “Alegría”, ordenados por el valor bicuadrado.

Tabla 16. Estimaciones de concordancia para cada evaluador en la categoría “Alegría” vs. todos los demás.

Evaluador	k_w total	
	Lineal	Bicuadrado
7	0.685	0.834
13	0.697	0.833
16	0.688	0.823
15	0.674	0.822
1	0.685	0.821
18	0.684	0.813
3	0.682	0.812
12	0.667	0.793
9	0.649	0.787
5	0.664	0.786
19	0.656	0.772
14	0.645	0.766
10	0.631	0.758
8	0.615	0.748
2	0.631	0.746
6	0.607	0.741
11	0.55	0.668
17	0.549	0.663
4	0.528	0.659

Con la idea de utilizar directamente las evaluaciones, la información proveniente de la proporción de concordancia observada, es conveniente hacer una depuración, se puede apreciar en la tabla 16, como los evaluadores 4, 11 y 17 tuvieron un nivel de concordancia

notablemente inferior con respecto a los otros evaluadores. En la tabla 17 se muestran los niveles de kappa logrados utilizando solo a los 10 evaluadores que mostraron mayor afinidad en sus evaluaciones, que son mayores en comparación con la Tabla 15.

Tabla 17. Estimaciones de kappa con 10 evaluadores.

Categoría	k_wtotal	
	Lineal	Bicuadrado
<i>Alegría</i>	0.726	0.860
<i>Enojo</i>	0.699	0.834
<i>Miedo</i>	0.743	0.872
<i>Repulsión</i>	0.705	0.837
<i>Sorpresa</i>	0.717	0.838
<i>Tristeza</i>	0.718	0.851

El alto nivel de concordancia nos da la confianza de utilizar la información de estos evaluadores para realizar la estimación estadística sobre la valoración de cada palabra, y con ello tener una proporción probabilística o del porcentaje con que se utiliza un término para denotar una emoción. Nótese que se usó el índice de kappa para elegir a los evaluadores más afines. A partir de este momento ya no usamos kappa. La matriz de evaluaciones tiene la forma mostrada en la tabla 18.

Tabla 18. Matriz con todos los datos de evaluaciones por cada emoción.

Evaluador	Palabra	Cat 1	...	Cat m
E ₁	P ₁	n ₁₁₁	...	n _{11m}
E ₁	P ₂	n ₁₂₁	...	n _{12m}
...
E ₁	P _N	n _{1N1}	...	n _{1Nm}
E ₂	P ₁	n ₂₁₁	...	n _{21m}
E ₂	P ₂	n ₂₂₁	...	n _{22m}
...
E _M	P _N	n _{MN1}	...	n _{MNm}

donde M es la cantidad de evaluadores, N es la cantidad de palabras y m es la cantidad de categorías de evaluación (nula, baja, media, alta). Nótese que los valores de esta tabla son 1 (positivos) o 0 (negativos) dependiendo si el evaluador eligió esta categoría o no.

El cálculo de la proporción de frecuencia de uso afectivo sobre cada palabra, se realizó de la siguiente manera: primero se calcula la cantidad de Evaluaciones Positivas (EP) para

cada palabra que recibió cada palabra en cada categoría (de nula hasta alta) de evaluación, y se vacían los resultados en una nueva matriz (EP). Para todos j (de 1 a N) y k (de 1 a m) se calcula el valor de cada celda de la matriz bidimensional EP:

$$EP_{jk} = \sum_{i=1}^M n_{ijk} \quad (13)$$

Para obtener los resultados en términos de porcentaje (PEP), es necesario multiplicar por cien y dividir entre la cantidad de evaluadores al contenido de cada celda de la matriz EP:

$$PEP_{jk} = \frac{100}{M} EP_{jk} \quad (14)$$

La tabla 19 es un segmento de las palabras de la categoría emocional “Alegría”, con el cálculo de porcentaje de probabilidad de uso de la palabra en un contexto relacionado con la alegría.

Tabla 19. Porcentajes de probabilidad de uso afectivo en palabras del diccionario “Alegría”.

PROBABILIDAD DE USO AFECTIVO					
#	Palabra	Nula[%]	Baja[%]	Media[%]	Alta[%]
1	<i>abundancia</i>	0	0	50	50
2	<i>acabalar</i>	40	0	60	0
3	<i>acallar</i>	50	40	10	0
4	<i>acatar</i>	50	40	10	0
5	<i>acción</i>	30	30	30	10
6	<i>aceptable</i>	0	20	80	0
7	<i>aceptación</i>	0	20	50	30
8	<i>acicate</i>	10	50	40	0
9	<i>aclamación</i>	10	10	10	70

Por ejemplo, en la palabra “abundancia”, en promedio el 50% de los evaluadores eligió la opción “Media”, lo que indica que esta palabra se utiliza en un contexto que denota alegría, casi con igual frecuencia que cuando se usa en un contexto donde no se denota dicha emoción; además, el otro 50% de los evaluadores eligió el valor “Alta”, lo cual indica que desde su punto de vista la palabra “abundancia” siempre o casi siempre se usa para denotar algo relacionado con la alegría.

En la palabra “acallar” el 50% eligió “Nula”, eso indica que jamás se usa para manifestar algo relacionado con la alegría; el 40% eligió “Baja”, indicando así que aunque sí puede usarse para denotar algo relacionado con la alegría, es un uso poco frecuente.

Como puede observarse, aunque los resultados presentados, ya pueden ser potencialmente útiles para sistemas automáticos de detección de categorías emocionales en texto a nivel frase, párrafo, texto o corpus, se propone otra característica de cada palabra que llamamos el factor de probabilidad de uso afectivo (FPA).

La propuesta considera que entre más cercanas estén las evaluaciones a la opción “Nula”, menor será el índice FPA, siendo ‘0’ el valor mínimo, y esto sucederá cuando **todos** los evaluadores hayan seleccionado la opción “NULA”. El valor máximo es ‘1’, y esto sucederá cuando **todos** los evaluadores hayan elegido la opción “Alta”. Para todos los valores intermedio y posibles combinaciones, se plantea una escala de pesos lineal:

$$w_k = \sum_{k=1}^m \frac{k - 1}{m - 1} \quad (15)$$

La estimación del factor de probabilidad de uso afectivo se define por:

$$FPA_j = \sum_{k=1}^m \frac{PEP_{jk} \cdot w_k}{100} \quad (16)$$

La tabla 20 es un segmento de la categoría emocional “Alegría”, donde se muestran los resultados obtenidos.

Tabla 20. Fragmento de los valores FPA para la categoría “Alegría”.

PROBABILIDAD DE USO AFECTIVO						
#	Palabra	Nula[%]	Baja[%]	Media[%]	Alta[%]	FPA
1	abundancia	0	0	50	50	0.830
2	acabalar	40	0	60	0	0.396
3	acallar	50	40	10	0	0.198
4	acatar	50	40	10	0	0.198
5	acción	30	30	30	10	0.397
6	aceptable	0	20	80	0	0.594
7	aceptación	0	20	50	30	0.696
8	acicate	10	50	40	0	0.429
9	aclamación	10	10	10	70	0.799

Para explicar lo que implica el FPA se analizan algunos resultados obtenidos:

1. Palabra “abundancia”, el 50% de los evaluadores cree que dicho término siempre denota algo relacionado con la alegría, el otro 50% cree que se usa en la misma proporción para

denotar algo relacionado o no con la alegría. Con estas evaluaciones se observa una alta correspondencia entre la palabra citada y la emoción alegría, de ahí su alto valor de FPA.

2. Palabra “acallar”, el 50% cree que este término no tiene relación alguna con la emoción alegría, 40% piensa que aunque si puede tener relación con la alegría, casi nunca se usa con esa finalidad, y solo el 10% lo evalúa con la categoría media. Se aprecia una pobre relación entre esta palabra y la emoción alegría, lo cual corresponde al bajo nivel FPA observado.

3. En las palabras “aceptable” y “aceptación”, ambas, tienen 0% en la opción “Nula” y 20% en la opción “Baja”; la diferencia es la distribución en las categorías superiores, en ellas se observa que “aceptación” tiene una mayor estimación hacia la parte “Alta”; por ello, el FPA es también mayor.

Por lo que se aprecia que a mayor FPA, mayor es la probabilidad de que la palabra sea usada con un sentido relacionado con la emoción con que está etiquetada. Queda a criterio libre establecer el uso del factor en sistemas de detección automáticos de emociones.

6. Conclusiones

El método presentado para la elaboración de diccionarios emocionales, se diferencia con respecto a los diccionarios de términos afectivos conocidos, principalmente por el hecho de que no fue concebido para indicar el nivel o “fuerza” que tiene una palabra con respecto a una emoción —por ejemplo, que tan positiva o negativa es una palabra—, sino que fue creado para indicar una estimación de la frecuencia con que una palabra se usa con un sentido relacionado a una emoción. Además se provee un factor de probabilidad de uso afectivo (FPA), el cual potencialmente se puede emplear en una etapa de un sistema automático de clasificación (redes neuronales, memorias asociativas, máquinas de soporte vectorial, etc.) como un elemento de ponderación de afectividad y mejorar así la precisión en el reconocimiento de categorías emocionales.

La etapa de análisis de concordancia tiene como finalidad dar un punto de referencia acerca de la calidad y validez de los términos seleccionados; también sirve como método para la “depuración” de los evaluadores. En este caso se utilizó el método de kappa ponderado propuesto por Cohen —adaptándolo para múltiples evaluadores—, sin embargo existen otras métricas que pueden utilizarse con la misma finalidad.

Es de mencionarse que existen algunas palabras que dependiendo de su contexto pueden relacionarse con más de una emoción, por esa razón se trabajó de manera independiente para cada categoría emocional propuesta, de ahí que algunas palabras existan en más de una categoría; se aclara que las palabras fueron evaluadas de manera independiente con respecto a su categoría emocional.

Adicionalmente, las palabras propuestas no son todas las existentes para denotar emociones, aunque aplicando el mismo método se pueden agregar más. También es necesario considerar que aunque se trate de hispanoparlantes, el sentido de las palabras cambia de una región a otra, y palabras que en alguno de sus sentidos puedan reflejar una emoción, en otra región es posible que no sea así, por lo que no se descarta la idea de agregar una etiqueta donde se mencione la región donde radiquen los evaluadores.

Por último, este recurso es una alternativa nueva y diferente a las utilizadas actualmente para resolver la identificación del contenido emocional en textos; además incrementa los limitados recursos léxicos existentes para el español en esta área, por lo que se pone a disposición de cualquier interesado el diccionario desde un sitio web³ o solicitándolo directamente a los autores por medio de sus cuentas de correo electrónico.

7. Bibliografía citada

BOLDRINI, Ester y otros, 2009: “EmotiBlog: an annotation scheme for emotion detection and analysis in non-traditional textual genres”, *Proceedings of the 2009 International Conference on Data Mining*, 491–497.

COHEN, Jacob, 1960: “A coefficient of agreement for nominal scales”, *Educational and Psychological Measurement* 20, 37–46.

COHEN, Jacob, 1968: “Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit”, *Psychological Bulletin* 70, 213–220.

CORTES, Édgar, 2008: *Comparación en la estimación del VO₂max a través de un monitor de frecuencia cardíaca Polar S810 y una prueba de esfuerzo maximal en banda sin fin según el protocolo de Balke, en deportistas universitarios entrenados en resistencia*

³ <http://www.cic.ipn.mx/~sidorov>

aeróbica en la ciudad de Bogotá, Tesis de Maestría en Epidemiología Clínica, Universidad Nacional de Colombia.

COWIE, Roddy y Randolph CORNELIUS, 2003: "Describing the emotional states that are expressed in speech", *Speech Communication Special Issue on Speech and Emotion* 40, 5–32.

EKMAN, Paul, 1972: "Universals and cultural differences in facial expressions of emotion", *Nebraska Symposium on Motivation* 19, 207–283.

ESULI, Andrea y Fabrizio SEBASTIANI, 2006: "SentiWordNet: A Publicly Available Lexical Resource for Opinion Mining", *Fifth international conference on Language Resources and Evaluation (LREC 2006)*, 417–422.

FLEISS, Joseph y Jacob COHEN, 1973: "The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability", *Educational and Psychological Measurement* 33, 613–619.

GREENBERG, Leslie, 2000: *Emociones: Una guía interna*, Bilbao: Descleé De Brouwer.

LANDIS, Richard y Gary KOCH, 1977: "The measurement of observer agreement for categorical data", *Biometrics* 33, 159–174.

MOLINER, María, 1996: *Diccionario de uso del español*, tercera edición, Madrid: Gredos.

MORENO, Antonio y otros, 2010a: "Sentitext: sistema de análisis de sentimiento para el español", *Procesamiento de Lenguaje Natural*, 45, 297–298.

MORENO, Antonio y otros, 2010b: "Análisis de Valoraciones de Usuario de Hoteles con Sentitext: un sistema de análisis de sentimiento independiente del dominio", *Procesamiento de Lenguaje Natural*, 45, 31–39.

MORENO, Antonio y otros, 2011: "Domain-neutral, Linguistically-motivated Sentiment Analysis: a performance evaluation", *XXVII Congreso de la Sociedad Española para el Procesamiento del Lenguaje Natural*, 361–369.

MORENO, Antonio y Chantal PÉREZ, 2013: "Lexicon-Based Sentiment Analysis of Twitter Messages in Spanish", *Procesamiento del Lenguaje Natural*, 50, 93–100.

ORTONY, Andrew y Terence TURNER, 1990: "What's basic about basic emotions?" *Psychological Review* 97, 315–331.

REDONDO, Jaime y otros, 2007: "The Spanish adaptation of ANEW (Affective Norms for English Words)". *Behavior Research Methods* 39, 600–605.

RUSELL, James, 1980: “A circumplex model of affect”, *Journal of Personality and Social Psychology* 39, 1161–1178.

SCHERER, Klaus y Paul EKMAN, 1984: “On the Nature and Function of Emotion: A Component Process Approach” en *Approaches to emotion*, Hillsdale New Jersey: Lawrence Erlbaum Associates, 293–317.

SMITH, Craig, 1989: “Dimensions of appraisal and physiological response in emotion”, *Journal of Personality and Social Psychology* 56(3), 339–353.

STONE, Phillip y otros, 1966: *The General Inquirer: A Computer Approach to Content Analysis*, Oxford England: The M.I.T. Press.

STRAPPARAVA, Carlo y Alessandro VALITUTTI, 2004: “WordNet-Affect: an affective extension of wordnet”, *Proceedings of the 4th International Conference on Language Resources and Evaluation*, 1083–1086.

STRAPPARAVA Carlo y Rada MIHALCEA, 2007: “SemEval-2007 Task 14: Affective Text”, *Proceedings of the 4th International Workshop on the Semantic Evaluations (SemEval 2007)*, Praga República Checa.

MARTÍN, Tamara y otros, 2010: “Opinion Polarity Detection - Using Word Sense Disambiguation to Determine the Polarity of Opinions”, *Proceedings of International Conference on Agents and Artificial Intelligence (1)*, 483–486.

WUNDT, Wilhelm, 1896: *Grundriss der Pshychologie*, Leipzig Alemania: Erlag von Wilhelm Engelmann.