

Encabezado: UN PROGRAMA DE COMPUTADORA QUE ANALIZA TEXTOS EN  
ESPAÑOL

*Publicado en la Revista Mexicana de Psicología, 24, 85-99*

La Psicología del Uso de las Palabras:

Un Programa de Computadora que Analiza Textos en Español

[The Psychology of Word Use: A Computer Program that Analyzes Texts in Spanish].

Nairán Ramírez-Esparza, James W. Pennebaker, Florencia Andrea García,

Universidad de Texas en Austin

y

Raquel Suriá

Universidad de Alicante, España

Nota del Autor: Para solicitar reproducciones de este artículo y del diccionario en español, así como también para solicitar análisis de textos en Español, por favor dirigirse a Nairán Ramírez-Esparza (nairan@mail.utexas.edu) o a James W. Pennebaker (pennebaker@psy.utexas.edu), en el Departamento de Psicología de la Universidad de Texas, Austin, TX 78712. Para solicitar el programa que analiza textos (LIWC) por favor dirigirse a la página electrónica <http://www.liwc.net/liwcspanol/>. Deseamos agradecer a Ángel Gómez, Gisela Sierra, Itziar Fernández, y Marina Beléndez, por su ayuda en revisar versiones anteriores de este manuscrito. La preparación de este trabajo fue posible gracias a la subvención otorgada por la “National Institutes of Health (MH52391)”.

### Resumen

El “Buscador Lingüístico y Contador de Palabras” (LIWC, por sus siglas en inglés, Pennebaker, Francis, & Booth, 2001) es un programa de computadora que analiza textos. Este programa calcula el porcentaje de palabras dentro de un texto de acuerdo a varias docenas de categorías. La fiabilidad de este programa ha sido demostrada ampliamente en el ámbito de la lengua inglesa. En esta investigación dos estudios se llevaron a cabo para analizar la equivalencia del programa en español al programa en inglés. En el Estudio 1 se presenta el procedimiento de traducción del LIWC del inglés al español, y se demuestra la equivalencia entre las categorías del LIWC en inglés y sus correspondientes categorías en la versión en español. En el Estudio 2 se muestra el uso del LIWC en inglés y en español al comparar el lenguaje utilizado por mujeres en foros de discusión en Internet de depresión y de cáncer de mama. Los resultados mostraron que las versiones correlacionan en la mayoría de las categorías. Asimismo, se encontró que las mujeres en foros de depresión utilizan distintas categorías de palabras que las mujeres en foros de cáncer de mama y estas diferencias son similares en foros de discusión en español e inglés. Se discuten las implicaciones de usar este programa dentro de la lengua española.

Palabras clave: Análisis de Texto, Programa de Computadora, LIWC, Internet, Foros, Depresión.

### Abstract

Linguistic Inquiry and Word Count (LIWC, Pennebaker, Francis, & Booth, 2001) is a computerized text analysis program. This program calculates the percentage of words within any given text according to several dozen categories. LIWC’s reliability has widely been demonstrated in the context of the English language. In this investigation, two studies were carried out to analyze the equivalence of the Spanish version of the program to the English version. Study 1 presents the translation procedure from English to Spanish, and tested equivalence between categories in the English LIWC and the corresponding categories in the Spanish version. Study 2 shows the use of both the English and Spanish LIWC version in comparing the language used by women in depression and breast cancer Internet chat rooms. The

results showed high correlations across English and Spanish LIWC categories. Furthermore, women in depression chat rooms used different categories of words than women in breast cancer chat rooms, and these differences were similar in chat rooms in English and in Spanish. The implications of using this program in the Spanish language are discussed.

Key words: Text Analysis, Computerized Program, LIWC, Internet, Chat Rooms, Depression

## La Psicología del Uso de las Palabras:

### Un Programa de Computadora que Analiza Textos en Español

Las palabras que usamos transmiten una gran cantidad de información acerca de quienes somos, a quién nos dirigimos, y las situaciones en las que estamos. Por ejemplo, las palabras que usa una persona pueden reflejar su personalidad, su estado de ánimo, sus motivaciones, su nivel socioeconómico, y una gran variedad de otros aspectos. El hecho de que las palabras indiquen quién es la persona y su estado emocional no es un concepto nuevo. Freud (1901) en sus escritos mencionaba sobre los actos fallidos de la lengua para indicar que a través de las palabras las personas transmiten sus sentimientos más profundos.

Una de las maneras más innovadoras y comprensivas de estudiar las palabras es mediante el uso de programas de computadora (Pennebaker, Mehl, & Niederhoffer, 2003). Estos programas típicamente están basados en un simple conteo de palabras y han sido utilizados en el área de las ciencias sociales y en psicología. Uno de estos programas, el “General Inquirer” (Stone, Dunphy, Smith, & Ogilvie, 1966), fue diseñado fundamentándose en teorías psicoanalíticas y uno de sus diccionarios fue creado para analizar el uso de palabras en historias dadas a las imágenes del TAT (El Test de Apercepción Temática). Otro programa, el “DICTION” (Hart, 1984), fue creado para analizar el lenguaje de discursos políticos y predecir si el uso de palabras en estos discursos influye en que un candidato sea electo o no. Un programa más reciente, el “Buscador Lingüístico y Contador de Palabras” (LIWC, por sus siglas en inglés, Pennebaker et al., 2001), fue diseñado para analizar las implicaciones psicológicas del uso de las palabras.

Desafortunadamente la mayor parte de las investigaciones que abordan el uso de las palabras y su impacto en aspectos psicológicos está principalmente limitado a la lengua inglesa. Con el propósito de abordar este problema, en esta investigación se introduce un programa de computadora capaz de analizar textos en español. Este programa es la versión en español del LIWC.

*La Importancia de Diseñar un Análisis de Texto en el Español*

Debido a que hoy en día vivimos un mundo de globalización, la investigación transcultural en el área de la psicología ha tomado un gran auge en los últimos años (Shiraev & Levy, 2004). De ahí que se ha emprendido la gran tarea de traducir cuestionarios en distintas lenguas para medir distintos factores psicológicos. Por ejemplo, en el área de la personalidad (e. g., Benet-Martínez & John, 1998), en el área del afecto (e. g., Joiner, Sandin, Chorot, Lostao, & Marquina, 1997), y en el área de la salud (Beck, Steer, & Brown, 1996, Páez, Martínez et al., 1999) entre otras (ver, Hofstede, 1983). Sin embargo, el uso de las palabras a través de las culturas ha sido explorado de manera muy limitada.

Es bien sabido que las culturas colectivistas son definidas por tener un “yo” interdependiente y las culturas individualistas por tener un “yo” independiente (Markus & Kitayama, 1991). Curiosamente, en una serie de estudios (Brewer & Gardner, 1996; Gardner, Gabriel, & Lee, 1999) se demostró que el uso de pronombres puede dar origen a distintos conceptos del “yo”. Más específicamente estos investigadores pidieron a un grupo personas que identificaran en un texto pronombres en primera persona del plural (e. g., nosotros) y que los circularan, y a otro grupo de personas les pidieron que identificaran pronombres en primera persona del singular (e. g., yo, mí) y que los

circularan. Después los autores midieron su yo colectivo y su yo individualista (e.g., mediante cuestionarios) y encontraron que las personas que identificaron pronombres en primera persona en plural tenían un auto-concepto más colectivo, a diferencia de los participantes que circularon los pronombres en primera persona singular.

Esta serie de investigaciones indica que las palabras (en este caso pronombres) tienen un efecto psicológico y cultural de manera que las personas pueden cambiar su referencia del “yo” ya sea a un “yo” colectivo (característico de países tales como México, España, o Japón) o un “yo” individualista (característico de países individualistas como Estados Unidos o Alemania). Debido a esta evidencia podemos inferir que el uso de las palabras tiene implicaciones psicológicas y culturales; sin embargo para poder explorar más esta idea se requiere de lo siguiente: a) que exista un programa de computadora que cuente las palabras y que tenga validez psicométrica en una lengua; b) que mida categorías lingüísticas gramaticales (tales como pronombres, adjetivos, preposiciones), así como categorías psicológicas y sociales (tales como emociones, familia, amigos); y c) que este programa sea traducido a otra(s) lenguas y tenga propiedades equivalentes.

El LIWC (Pennebaker et al., 2001) es un programa que analiza textos y que cubre los puntos ‘a’ y ‘b’ antes mencionados. Por ejemplo, este programa cuenta el porcentaje de palabras que caen en un gran espectro de categorías lingüísticas y psicológicas y ha demostrado tener propiedades psicométricas en inglés (e. g., Alpers et al., 2005; Mehl & Pennebaker, 2003; Pennebaker & King, 1999). En este estudio se abordará el punto ‘c’, es decir, se introducirá la versión al español del LIWC y se mostrará su aplicación en el área de la psicología. A continuación se da una descripción más detallada del LIWC en

inglés, después se presentan ejemplos de cómo el LIWC se ha utilizado en el área de la Psicología y, finalmente, se presentan dos estudios para introducir el LIWC en español.

### **Buscador Lingüístico y Contador de Palabras (LIWC por sus Siglas en Inglés)**

El LIWC (Pennebaker et al., 2001) se desarrolló dentro del contexto del trabajo de Pennebaker sobre escritura emocional (e. g., Páez, Velasco, & González, 1999; Pennebaker & Francis, 1996). Pennebaker y Beall (1986) encontraron en una serie de estudios que las personas mejoraban su salud después de escribir sobre experiencias traumáticas. Por lo tanto Pennebaker et al. (2001) diseñaron el LIWC para descubrir cuáles son las características de escribir experiencias negativas que pueden predecir una mejor salud en un futuro. Más recientemente, la aplicación del LIWC se ha extendido para analizar el uso del lenguaje en fuentes de texto, literatura, narraciones personales, y conversaciones cotidianas (Pennebaker & Graybeal, 2001).

#### *Descripción General del LIWC*

Las Categorías. El LIWC contiene 72 dimensiones lingüísticas, las cuales fueron creadas a partir de conceptos psicológicos y con la ayuda de jueces. Estas dimensiones incluyen categorías de lenguaje estándar (e. g., artículos, preposiciones, pronombres, incluyendo primera persona en singular, primera persona en plural etc.), procesos psicológicos (e. g., categorías de emociones positivas y negativas, procesos cognitivos), palabras relacionadas a la relatividad (e. g., tiempo, tiempo del verbo, moción, espacio), y dimensiones tradicionales de contenido (e. g., sexo, muerte, casa, ocupación). Las categorías del LIWC están organizadas jerárquicamente, por ejemplo, todas las palabras relacionadas con la categoría ‘enojo’ también se organizan en las categorías de emociones negativas o palabras afectivas.

Funcionamiento. El LIWC utiliza una estrategia de conteo de palabras donde busca sobre 2,300 palabras y raíces de palabras dentro de un texto dado. Cada palabra y/o raíces de palabra es parte de una o más categorías o subcategorías de palabras. Por ejemplo, la palabra ‘llorar’ caería dentro de las siguientes cuatro categorías: ‘tristeza’, ‘emoción negativa’, ‘afecto en general’, y ‘tiempo pasado’. Por lo tanto, si esta palabra es encontrada en un texto, cada una de las escalas de las mencionadas categorías se incrementará. Por otro lado, el programa toma en cuenta el porcentaje total de palabras. Por ejemplo, si el número resultante es 9.25 en la categoría “pronombres”, esto significa que el 9.25% del total de palabras de un texto fueron pronombres. Las únicas categorías que no reflejan porcentajes son [CP] que se refiere al número de palabras en el texto, [PPO], es decir, el número de palabras por oración, y [SIGNOS?], número de oraciones que terminan con signo de interrogación.

Para que el LIWC funcione se deben de usar diccionarios. Lo único que se tiene que hacer es pedirle al programa que abra la opción de “descargar un nuevo diccionario” del Menú de Diccionarios. Los diccionarios externos deben estar en archivos de texto y se deben archivar con la terminación *dic* (e. g., español.dic). Por ejemplo, en un fólдер uno puede tener el diccionario en español, el diccionario en inglés, el diccionario en alemán. El interesado simplemente selecciona el que desee usar en ese momento. Asimismo, uno puede crear sus propios diccionarios. Por ejemplo, si el investigador está interesado en saber cuántas palabras relacionadas con las selecciones en México se han usado en foros de discusión, el interesado podría crear un diccionario con palabras claves (e.g., Calderón, Lopez Obrador, manifestaciones, etc.). En la Tabla 1 se muestra un ejemplo de cómo crear un diccionario con referencia a uno mismo. Como se puede ver en

la Tabla 1, los números indican las categorías a las cuales cada palabra es asignada. En este caso se crearon 10 categorías. Por ejemplo, la palabra “mi” corresponde a la categoría 2 (que indica solamente la palabra “mi”), la categoría 7 (que indica primera persona singular), y la categoría 10 (que cuenta todas las palabras que se refieren al yo). En el LIWC las raíces de palabra son palabras parciales seguidas por un asterisco. Por lo tanto un asterisco en el diccionario al final de una parte de palabra indica al LIWC incluir las letras subsecuentes. En el diccionario ejemplificado en la Tabla 1, la palabra “mí\*” indica a LIWC incluir palabras tales como “mío” y “mía”.

#### *Propiedades Psicométricas del LIWC y su Utilidad en el Área de la Psicología*

Características Psicométricas. El LIWC ha demostrado tener estabilidad a través del tiempo y consistencia a través de contextos. Pennebaker y King (1999) analizaron un gran número de textos tomados de diarios, trabajos de escuela, resúmenes de revistas científicas, y se encontró una consistencia alta a través del tiempo (i.e., a través de textos escritos por la misma persona en diarios), tema (i.e., escritos de la misma persona con temas diferentes), y el origen del texto (i.e., el mismo tipo de texto escrito por diferentes personas). En un estudio de campo Mehl y Pennebaker (2003) recolectaron conversaciones de estudiantes en su vida cotidiana utilizando una pequeña grabadora, y también encontraron una alta consistencia en el uso de las palabras a través del tiempo y el tipo de contexto social. Recientemente Alpers et al. (2005) encontraron que la codificación de textos en distintas categorías hecha por humanos, correlacionó significativamente con la codificación hecha por el LIWC.

Variabes Demográficas. Pennebaker y Stone (2003) demostraron una relación entre edad y el uso de las palabras. Específicamente, mostraron que a mayor edad, los

individuos utilizan más palabras emotivas positivas, menos palabras negativas emotivas, menos referencias a la primera persona singular, más uso de palabras en tiempo futuro, y menos uso de palabras en tiempo pasado. Por otro lado, se ha encontrado que hay diferencias en el uso de las palabras entre hombres y mujeres (ver Mehl & Pennebaker, 2003; Pennebaker & King, 1999). Por ejemplo, al escribir, las mujeres utilizan más palabras en primera persona del singular que los hombres (Pennebaker & King, 1999) y en las conversaciones cotidianas los hombres utilizan más groserías y artículos, y las mujeres utilizan más palabras o expresiones de “relleno” (e. g., “este”, “mmm”) y más palabras que indican discrepancia (e. g., “debería”, “podría”, “debo”) (Mehl & Pennebaker, 2003).

Variables psicológicas y sociales. El uso de las palabras también está relacionado con la personalidad. Por ejemplo, se ha encontrado que el Neuroticismo se relaciona con una mayor frecuencia del uso de palabras en primera persona, así como un mayor uso de palabras con valencia negativa y menos uso de palabras con valencia positiva (ver Pennebaker & King, 1999 para una descripción más detallada). Por otro lado, en otro estudio se encontró que las personas que dicen la verdad utilizan más pronombres en primera persona que los individuos que mienten (Newman, Pennebaker, Berry, & Richards, 2003). También se ha mostrado que después de experimentar un trauma colectivo (e. g., El 11 de Septiembre, la muerte de la Princesa Diana, etc.), las personas experimentan “un descenso” o “una reducción” en el uso del pronombre “yo” e incrementa el uso del pronombre “nosotros” (Cohn, Mehl, & Pennebaker, 2004; Stone & Pennebaker, 2002).

Hasta este punto se ha presentado que una de las maneras más comprensivas de analizar el uso de las palabras es mediante la utilización de un programa de computadora. Sin embargo, se ha hecho notar que estos programas y su aplicación en el área de la psicología solamente se ha realizado en los países de habla inglesa. De ahí la necesidad de crear un programa que pueda usarse en las culturas hispanas y que mida aspectos tanto lingüísticos como psicológicos. Se ha introducido el LIWC, cuyas propiedades psicométricas, su aplicación en el área de la psicología, y su fácil manejo y flexibilidad, lo hace el programa ideal para traducirlo al español.

A continuación se presentarán dos estudios donde se introduce el LIWC en español. En el primer estudio se mostrará como el LIWC en inglés fue traducido al español y se pone a prueba la equivalencia de ambas versiones. Es decir, se observa si las categorías del LIWC en inglés son similares a las categorías del LIWC en español. En el segundo estudio, se aplica el LIWC en inglés y español al comparar el lenguaje usado por mujeres en foros de discusión en Internet de depresión y de cáncer de mama.

### **Estudio 1: El Desarrollo del LIWC en Español**

El propósito de este estudio fue traducir el diccionario en inglés del LIWC al español, así como observar si las categorías del LIWC en español son equivalentes a las categorías de la versión en inglés. A continuación se describe más detalladamente cada uno de los pasos.

#### *Traducción del LIWC del Inglés al Español*

La ardua tarea de traducir el LIWC del inglés al español duró aproximadamente 1 año y medio. Esta tarea fue hecha—bajo la coordinación de la primera autora de este manuscrito—por una asistente bilingüe estudiante de psicología y que trabajaba tiempo

parcial en el departamento de psicología social en la Universidad de Texas en Austin. La labor consistió en traducir con exactitud las palabras del inglés al español y dejarlas en las categorías que originalmente aparecen en el LIWC en inglés. Por ejemplo, el pronombre “I” en inglés se tradujo como “Yo” en español y se mantuvo en sus categorías correspondientes: 1 (Pronombre), 2 (Primera persona del Singular), y 4 (Total primera persona). La fase de traducción se hizo con la ayuda de un diccionarios inglés/español, inglés/inglés (e. g., “Basic Dictionary of American English”), español/español, y diccionarios en línea (e. g., [www.foreignword.com](http://www.foreignword.com) y [www.babylon.com](http://www.babylon.com)). Es importante destacar que algunas palabras en español no se dejaron en las mismas categorías que en inglés. Esto con el objetivo de transmitir ideas que son congruentes en un idioma, pero no en el otro. Por ejemplo, la palabra “blue” en inglés significa azul y estar triste, pero en español la palabra “azul” significa simplemente azul. Igualmente es importante mencionar que se agregaron palabras en el diccionario español que no existían en el diccionario en inglés. Esto con el objetivo de mejorar el diccionario en español, en comparación con el diccionario en inglés. Por ejemplo, la palabra “rimbombante” no existe en el diccionario en inglés, pero sí en la versión en español. El proceso de traducir el diccionario de inglés al español, es comparable con el proceso de traducir un ensayo del inglés al español. Es decir, para transmitir la misma idea en otro idioma, se requiere que se agreguen palabras o se quiten palabras.

El segundo paso fue conjugar los verbos, tanto en la forma singular como en el plural, así como en los tiempos presente, pasado y futuro (ya que en el LIWC en inglés solamente existen categorías para estos tres tiempos). El tercer paso consistió en confirmar que las categorías originalmente asignadas en inglés tuvieran sentido en

español incluyendo las modificaciones pertinentes. Así, por ejemplo, el verbo “take” en inglés, se puede traducir al español como tomar, agarrar, coger.

El paso final consistió en una revisión del LIWC en español por tres jueces. La juez 1 era mexicana, estudiante de postgrado en lingüística en la universidad de Texas en Austin. La juez 2 fue una estudiante de licenciatura bilingüe originaria de México a quien se le pagó durante un semestre para revisar el diccionario en español. La juez 3 fue una española, doctora en Psicología, que se dedicó a limpiar y pulir la última versión del LIWC. En general, los jueces revisaron cada una de las palabras, si estaban bien escritas, conjugadas, y en la categoría apropiada. Por ejemplo, la segunda palabra del diccionario en español es “abajo”, el juez veía la palabra, si estaba bien escrito y si estaba organizada en sus categorías correspondientes. En este caso “abajo” quedó en la categoría 10 (preposición), en la categoría 41 (Espacio) y en la categoría 43 (Abajo), y así para cada palabra. Este último paso se desarrolló durante un año aproximadamente. A partir de las correcciones de los jueces se elaboró una versión final del LIWC en español.

Versión Final del LIWC en Español. La versión final del diccionario en español cuenta con un total de 7,515 palabras y raíces de palabras, y con el mismo número de categorías y subcategorías que el LIWC en inglés (i.e., 72, ver Tabla 2 donde se mencionan las categorías y se dan ejemplos).<sup>1</sup>

El diccionario en español tiene significativamente más palabras que la versión en inglés debido a que en español existen más conjugaciones verbales, se consideran las palabras en femenino y masculino, palabras acentuadas y no acentuadas (e. g., camino, caminó), y sinónimos de verbos que son muy usados en esta lengua pero no en inglés (e. g., “take” en inglés es el verbo más usado para decir “agarrar algo”, pero en español es

común utilizarlo tanto para “tomar”, “agarrar”, o “coger”). A pesar de que el diccionario en inglés también tome en cuenta variaciones de palabras, en el diccionario en español resultó necesario aumentar el número de palabras para transmitir la misma idea. Por ejemplo, en el diccionario en inglés si uno quisiera que todas las palabras que se refieren a “amor” se tomaran en cuenta en una categoría, se pondría solamente lov\*. De esta manera se incluirán las palabras love, loving, loved, lovely, lover, etc. Sin embargo, si en español se quisiera que todas las palabras relacionadas con amor se incluyeran en tal categoría, se tendría que escribir amor, amor\*, amo, ama\*, amé, amó.

#### *Equivalencia del LIWC en Inglés y Español*

Con el objetivo de comprobar la equivalencia de ambos diccionarios, se analizaron textos de los que se dispusiera de traducciones en inglés y español. Esto dio la opción de analizar textos que tuvieran la misma información en ambos idiomas. Un total de 83 textos se obtuvieron de Internet con sus respectivas traducciones. Los temas variaron desde cuestiones políticas, noticias, salud, bibliografías, propaganda, anuncios, canciones, reflexiones, recetas, poemas, y plegarias, entre otros.

Los textos en inglés se analizaron con el LIWC en inglés y los textos en español se analizaron con el LIWC en español. Se guardaron los datos en un paquete estadístico y con el propósito de ver la equivalencia entre ambos diccionarios se hicieron los siguientes análisis. Primero se hicieron correlaciones de Pearson para estudiar si cada una de las categorías en inglés correlacionaba con las correspondientes categorías en español. Por ejemplo, se observó si la categoría pronombres en inglés correlacionaba con la categoría pronombres en español y así sucesivamente para las 72 categorías. Cabe recordar que existen el mismo número de categorías en el LIWC inglés que en el LIWC en español,

por lo tanto fue posible correlacionar uno a uno cada una de las categorías. Segundo, se hicieron pruebas *t*-dependientes (o relacionadas) para observar si las medias obtenidas en cada una de las categorías en inglés eran similares a la medias obtenidas en las correspondientes categorías en español. Por ejemplo, se observó si la media de porcentaje de palabras correspondiente a la categoría de pronombres en el LIWC en inglés era similar a la media de porcentaje de palabras correspondiente a la misma categoría en español.

Las correlaciones observan si la media de porcentaje de palabras obtenidas en las categorías del LIWC en inglés va en la misma dirección que la media de porcentaje de palabras obtenidas en las categorías del LIWC en español. Si la correlación es alta quiere decir que al analizar los textos, el LIWC en inglés y el LIWC en español contabilizaron las palabras de manera similar. Por ende, esto demuestra que las versiones capturan el mismo contenido dentro de un texto. Por otro lado, se utilizaron pruebas *t*-pareadas para observar ver si las medias obtenidas con el LIWC en inglés en cada categoría son iguales a las medias en sus correspondientes categorías en español. A pesar de que las correlaciones sean altas, no quiere decir que las medias sean exactamente iguales en cada versión. Por lo tanto al hacer pruebas *t*-pareadas se pone a prueba que tan similares son las medias. Si las medias son iguales entonces indicaría que las versiones son equivalentes. Si no son iguales quiere decir que una versión del programa cuenta más/menos palabras en ese texto. Dado que el LIWC en español tiene más palabras y se agregaron algunas palabras en el proceso de traducción que no existían en el LIWC en inglés, es probable que el LIWC en español sea capaz de capturar más palabras.

Resultados. La media del número de palabras contabilizadas por el LIWC en inglés en los textos en inglés fue de 642.34 ( $DS = 487.65$ ), y la media del número de palabras contabilizadas por el LIWC en español en los textos en español fue de 661.33 ( $DS = 480.72$ ). Asimismo, el LIWC en inglés contabilizó el 61.85% de las palabras en los textos y el LIWC en español contabilizó el 66.32% de las palabras. En la Tabla 2 se muestran los resultados de los análisis para cada una de las categorías. Específicamente, en la primera columna se incluyen las correlaciones entre las categorías en inglés y sus correspondientes categorías español. Como se puede observar, la mayoría de las categorías mostraron correlaciones significativamente altas, lo que indica que las versiones en inglés y en español son equivalentes. En otras palabras, si el LIWC en inglés contabilizaba las palabras en los escritos en inglés, también el LIWC en español tendió a contabilizar las mismas palabras en los escritos en español.

A partir de la siguiente columna se indican la media y la desviación estándar para cada opción en español e inglés, así como también el valor de la prueba  $t$  y su nivel significativo. Como se puede observar en la Tabla 2, se encontró que en 31 categorías las medias difieren entre ambas versiones ( $p < .01$ ). En el resto de las categorías (i.e., 41) se encontró que las medias no difieren en ambas versiones.

### *Discusión*

El propósito de este estudio fue llevar a cabo la adaptación del LIWC del inglés al español, así como demostrar la equivalencia de ambas versiones. Los resultados mostraron que, como se esperaba, el LIWC en español contó aproximadamente 4% más de las palabras contadas por el LIWC en inglés. Asimismo, se encontró que las versiones correlacionaron en la mayoría de las categorías, indicando una correspondencia adecuada

entre los diccionarios. Sin embargo, al analizar la diferencia de medias (i.e., las pruebas  $t$ ) se mostró que no todas las categorías son comparables entre las versiones.

¿Por qué no son comparables las versiones en algunas categorías? El hecho de que el LIWC en español difiriera del LIWC en inglés en algunas categorías no es un hallazgo inesperado. En primera porque el LIWC en español cuenta más palabras y en segunda, por los bien conocidos problemas de traducción. Por ejemplo, para transmitir una idea similar en otra lengua muchas veces se tienen que agregar o quitar palabras. Además la naturaleza del diccionario cambia a medida que la naturaleza del lenguaje cambia. Por ejemplo, la gramática del español es diferente a la gramática en inglés. Finalmente, para decir una idea en un lenguaje no se pueden usar la misma palabra en otro lenguaje (e.g., “blue” en inglés significa estar triste, “azul” en español significa azul). Todos estos aspectos jugaron un rol al momento de traducir el LIWC del inglés al español. Por ende, no es sorprendente que no se encontraran medias similares en todas las categorías. A continuación discutimos otras cuestiones más específicas de porque no se encontraron similitudes.

Una de las categorías que difirió significativamente fue “YO” (o primera persona del singular). Esto se puede deber a que en inglés, al referirse a la primera persona en singular, siempre se incluye el pronombre “yo”. Sin embargo, en español se omite. Esto explica porque otras categorías de pronombres no resultaron ser iguales en ambas versiones.

Otro grupo de categorías que difirieron fueron aquellas que incluyen verbos en los distintos tiempos (e. g., presente, pasado, futuro). Como se mencionó anteriormente, en español existen más conjugaciones de verbos que en inglés. Por lo tanto, esto indica que

probablemente, en un texto en español se utilizan más opciones de verbo para indicar un tiempo.

Por último, se puede observar que otro grupo de palabras que difirieron en ambas versiones fue aquellas categorías que incluye personas (e. g., amigos, familia, hermanos). Esto se puede deber a que en español se usan palabras para indicar femenino o masculino que en inglés (e. g., primo-prima, amigo-amiga, adulto-adulta, humano-humana, etc.). Es importante hacer notar que las categorías que resultaron ser similares a través de las adaptaciones de diccionarios, fueron aquellas que incluyen más sustantivos que verbos o que variaciones de singular y plural (e.g., Deportes, Hogar, Televisión, Música, etc). Es más fácil traducir este tipo de categorías directamente al español, porque nada más se traduce literal el sustantivo, pero no se agregaron variaciones de verbo o de género.

Es importante señalar que no importa que algunas categorías no sean equivalentes, esto no hace al LIWC en español menos confiable para medir aspectos psicológicos que el LIWC en inglés. Sin embargo, este es un problema cuando se quiere hacer un estudio transcultural. Si se quisiera investigar si las personas de habla hispana usan más palabras con connotación positiva en sus conversaciones diarias que las personas de habla inglesa, el LIWC no sería una herramienta eficaz. Porque el LIWC en español cuenta más palabras con respecto a emociones positivas que el LIWC en inglés.

#### *Limitaciones del LIWC*

El LIWC en español tiene las mismas limitaciones que el LIWC en inglés. Primero, las categorías están arbitrariamente diseñadas por investigadores y la organización de las palabras en sus respectivas categorías está hecha por un número reducido de jueces. Sin embargo, cabe señalar que el LIWC en inglés ha estado

mejorándose en los últimos 5 años y aún así cuenta aproximadamente el mismo número de palabras que la versión español (i.e., en este estudio 61.85% fueron contadas por la versión en inglés y 66.32% fueron contadas por la versión en español). Esto demuestra que la primera versión del LIWC en español al igual que la versión en inglés, cuenta las palabras más comúnmente usadas en un texto, pero no es perfecto. Lo ideal sería diseñar un diccionario que cuente más porcentaje de palabras en un texto dado. Futuras investigaciones deberían de tomar esto en cuenta, dado que es posible que algunas de las palabras no contadas también sean relevantes para entender mecanismos psicológicos.

La segunda limitación es que el LIWC se basa en el simple conteo de palabras y por lo tanto se desconoce el contexto en que se mencionaron dichas palabras. Sin embargo, cabe mencionar que el LIWC es tan acertado como codificaciones hechas por humanos. En un estudio Alpers et al. (2004) le pidió a codificadores humanos que evaluaran 100 textos con respecto a categorías tomadas del LIWC. Específicamente el codificador humano tenía que leer el texto e indicar del 0 (para nada) a 3 (Mucho) que tanto el texto engloba cada una de las categorías seleccionadas del LIWC. Los resultados mostraron que los codificadores humanos fueron tan acertados como el LIWC. Esta característica hace al LIWC una herramienta útil y práctica para analizar miles y miles de textos en cuestión de segundos donde el investigador puede ahorrar mucho tiempo y dinero.

Una tercera limitación del LIWC en inglés y el español es que no tiene categorías con respecto a todas las conjugaciones verbales. Por ejemplo en inglés no existe una categoría separada para verbos en presente participio (“writing”) o en pasado participio (“written”). Asimismo el LIWC en español no incluye una categoría en pospretérito

(escribiría) o en pretérito subjuntivo (escribiera o escribiese) y así sucesivamente para todas las conjugaciones en español. Esto es porque el LIWC fue creado para observar variables psicológicas más que lingüísticas. Sin embargo, cabe señalar que si el investigador está interesado en estas categorías, estas pueden ser agregadas al diccionario.

Una limitación del LIWC en español es que está en sus inicios y no se han hecho muchos estudios para analizar el lenguaje en distintos tipos de textos tal como se ha hecho con el LIWC en inglés (ver Pennebaker et al., 2001). Por lo tanto futuras investigaciones deben de tomar en cuenta esta variable y estudiar cómo el LIWC en español analiza textos que tienen distintas longitudes y distintos temas.

### **Estudio 2: La Aplicación del LIWC en Español en el Área de la Psicología**

Este estudio se llevó a cabo con el propósito de ver la aplicación del LIWC en español en el área de la Psicología. Específicamente se investigó un tema que ha sido ampliamente estudiado con el LIWC en inglés: la depresión. Así, se analizó las palabras utilizadas por mujeres en foros de discusión de depresión y las comparamos con un grupo control. A continuación se revisan algunas investigaciones previas que se han llevado a cabo en Estados Unidos y después se introduce los detalles de una investigación realizada con el LIWC en inglés y en español.

#### *Las Teorías de la Depresión*

Las teorías y perspectivas que explican los mecanismos asociados a la depresión son abundantes. Sin embargo, en este estudio nos centramos en tres. La primera menciona que las personas deprimidas tienden a pensar en una gran medida en ellos mismos. Por ejemplo, Pyszczynski y Greenberg (1987) han especulado que las personas

deprimidas después de que han perdido su fuente de autovaloración, se ven sumergidos en tratar de ganar lo que han perdido, por lo tanto tienden a enfocarse en sí mismos. De acuerdo a esta teoría se podría argumentar que las personas que están deprimidas van a tender a utilizar más palabras asociadas a centrarse en sí mismos que las personas que no están deprimidas.

Beck (1967), en su teoría cognitiva de la depresión, postula que las personas que están deprimidas tienden a verse así mismos y al mundo en términos negativos. De acuerdo a esta teoría se puede especular que las personas que están deprimidas van a tender a usar más palabras con valencia negativa (y menos con valencia positiva) que las personas que no están deprimidas.

Una tercera perspectiva es el modelo del suicidio llamado integración social/desconexión de Durkheim (1951). Este modelo establece que una de las claves que llevan al suicidio es la percepción de que uno mismo no está integrado en la sociedad, o en otras palabras, que está desconectado de la vida social. De acuerdo a esta teoría, uno puede especular que las personas que están deprimidas van a tender a usar menos palabras asociadas a procesos sociales (e. g., menos palabras en primera persona del plural).

#### *Estudios Hechos con el LIWC en Inglés*

En un estudio, Stirman y Pennebaker (2001) utilizaron el LIWC en inglés y compararon el uso del lenguaje en nueve poetas que se suicidaron y nueve poetas que no lo hicieron (aunque en este estudio no se midió depresión, se puede inferir que los poetas que se suicidaron estaban más deprimidos que los poetas que no se suicidaron). Se encontró que los poetas que se habían suicidado utilizaban más palabras en primera

persona del singular (e. g., yo, mi, mío) y menos palabras en primera persona del plural (e. g., nosotros, nuestro). Asimismo, los poetas suicidas tendían a utilizar menos palabras referentes a otros. Un hallazgo poco inesperado fue que los grupos no difirieron en el uso de palabras negativas o positivas.

En otro estudio, Rude, Gortner, y Pennebaker (2004) utilizaron el LIWC para analizar ensayos libres (i.e., ensayos sobre emociones y pensamientos más profundos de estar en la universidad) escritos por estudiantes deprimidos y estudiantes no deprimidos. En este estudio los niveles de depresión se midieron con respecto a auto-evaluaciones dadas al Inventario de Depresión de Beck (Beck, Ward, Mendelson, Mock, & Erbaugh, 1961). De acuerdo a lo esperado con respecto a la teoría de Pyszczynski y Greenberg (1987), los estudiantes deprimidos utilizaron el pronombre “yo” con mayor frecuencia que los no deprimidos. Con respecto a lo esperado según la teoría de Beck (1967), se encontró que los estudiantes deprimidos utilizaron más palabras con connotación negativa y marginalmente (i.e.,  $p = .08$ ) menos palabras con valencia positiva que los estudiantes no deprimidos. Finalmente, con respecto a lo esperado de acuerdo a la teoría de Durkheim (1951), no se encontraron diferencias en el uso de palabras asociadas a procesos sociales. Además, no se pudo analizar el uso de palabras en primera persona en plural porque la frecuencia del uso de estas palabras fue muy baja.

A partir de los hallazgos resultantes de estos estudios surgen las siguientes preguntas de investigación: ¿las personas deprimidas y no deprimidas de habla hispana mostrarán diferencias en el uso del lenguaje?<sup>2</sup> Asimismo, si existen diferencias, ¿serán similares a las encontradas con el uso del inglés? Como no existen estudios que se hallan hecho en esta área, no se puede especular exactamente que diferencias se vayan a

encontrar. Sin embargo, dado que las personas de habla hispana tienden a pertenecer a una cultura colectivista y las personas de habla inglesa tienden a pertenecer a una cultura individualista (Markus & Kitayama, 1991), se puede inferir que esto va a afectar los procesos psicológicos asociados a la depresión. Por ejemplo, probablemente las personas de habla hispana hablan de procesos sociales no importa si están deprimidos o no. O tal vez tienden a ser más optimistas no importa si están deprimidos o no. Por tanto, resulta interesante observar si los mecanismos de depresión, antes mencionados, también están presentes en la cultura de habla hispana.

Una de las maneras más prácticas y convenientes de tener acceso a estas poblaciones es mediante el uso de Internet. El hecho de que en los últimos años hayan surgido foros de discusión en línea da la oportunidad de observar y analizar de una manera discreta la forma en que las personas hablan acerca de situaciones específicas (Davison, Pennebaker, & Dickerson, 2000). En los foros de discusión las personas típicamente intercambian sus experiencias en línea con respecto a un tema. En este caso se utilizó foros de discusión de depresión y foros de discusión de cáncer. Específicamente se seleccionó mujeres que hayan escrito sobre su depresión en foros de discusión de depresión y se seleccionó mujeres que hayan escrito sobre su experiencia de tener cáncer de mama (grupo control). Se utilizó este grupo como control porque en los foros de cáncer de mama, al igual que los foros de depresión, las mujeres dan su testimonio de vivir con un padecimiento, y expresan sus emociones, sus sentimientos, y platican sobre como su padecimiento ha afectado su vida en cuestiones sociales y psicológicas. De esta manera ambos grupos son comparables.

### *Método*

Participantes. Un total de 160 textos se obtuvieron de los foros de discusión en español. De esos 80 textos fueron de mujeres de foros de depresión (<http://www.forodede depresion.com.ar/>) y 80 de foros de cáncer de mama (<http://www.todocancer.com/>). Por otro lado, un total de 160 textos se obtuvieron de foros de discusión en inglés. De esos 80 fueron de mujeres de foros de depresión (<http://www.dailystrength.org/>) y 80 fueron de mujeres de cáncer de mama (<http://www.cancerforums.net/>). Es importante hacer notar que un texto corresponde a una mujer. Es decir, no se obtuvieron múltiples textos por cada participante. No se logró conocer otras características personales de los participantes (i.e., su edad y de donde eran).

Procedimiento. Se seleccionaron muestras de textos de foros donde las personas escribieran un mínimo de 100 palabras. Por ejemplo, se acudía a un foro de discusión de depresión donde una persona se presentaba al grupo y hablaba sobre su depresión, se tomaba el texto, se corregían los errores de ortografía y se guardaba en formato de texto (e. g., de sujeto1.txt), y así sucesivamente. Lo mismo se hizo para los foros de discusión de cáncer de mama. Los análisis lingüísticos de los textos en español se analizaron con el LIWC en español y los análisis lingüísticos de los textos en inglés se analizaron con el LIWC en inglés. Se seleccionaron categorías relacionadas con las teorías de depresión antes mencionadas y que hubieran sido utilizadas en otras investigaciones (e. g., Rude et al., 2004; Stirman & Pennebaker, 2001). Estas categorías se refieren al uso de palabras en primera persona singular (e. g., yo, mí, mi), el uso de palabras con valencia positiva (e. g., contento, bonito, bueno, amor) y valencia negativa (e. g., odio, enemigo, nervioso, miedo), el uso de palabras en primera persona del plural (e. g., nosotros, nuestro) y

palabras sociales (e. g., mención de amigos, familia o comunicación). También ver Tabla 3 donde se presentan las categorías seleccionadas.

### *Resultados y Discusión*

El objetivo de esta investigación fue analizar si las personas de habla inglesa e hispana que participan en foros de depresión utilizan diferentes palabras en comparación con las personas que participan en foros de cáncer de mama. Para lograr este objetivo realizamos un total de 5 pruebas *t*-independientes para el grupo de habla hispana y para el grupo de habla inglesa. No hicimos comparaciones estadísticas por lenguaje debido a que, como se mencionó, en el Estudio 1 algunas categorías no son similares a través de los lenguajes. Es decir, si llegáramos a encontrar que las personas de habla inglesa deprimidas utilizan más palabras en primera persona singular, esto sería por el sesgo inherente de que el LIWC en inglés contabiliza más palabras en esta categoría y no porque las personas que participan en foros de depresión de habla inglesa las utilizan con mayor frecuencia que las personas que participan en foros de depresión de habla hispana. Sin embargo, al hacer pruebas *t*-independientes para cada grupo podemos analizar si las medias se comportan de la misma manera en cada lenguaje. Por ejemplo, podemos observar en los foros de habla inglesa si las mujeres en foros de depresión utilizan más palabras en primera persona singular que las mujeres en foros de cáncer de mama, y si este efecto es similar dentro de los foros de habla hispana.

*Características de los textos.* La media de palabras contadas por el LIWC en los foros de discusión en español fue de 263.78 (116.13) para el de depresión y 253.26 (121.96) para el de cáncer de mama. Por otro lado, la media de palabras contadas por el LIWC en los foros de discusión en inglés fue de 229.00 (115.54) para el de depresión y

246.88 (105.55) para el de cáncer de mama. No hubo diferencias significativas en el número de palabras contadas a través de los foros en inglés y español. Esto indica que las diferencias en las categorías de palabras usadas a través de los foros no se deben al número de palabras usadas.

La Tabla 3 muestra la media y la desviación estándar tanto para el grupo de foros de depresión y el grupo de foros de cáncer de mama en español y en inglés. Las dos últimas columnas muestran los valores  $t$  para los efectos principales para cada grupo (i.e., la prueba  $t$  para las diferencias encontradas en los foros de discusión en español al comparar el foro de depresión con el foro de cáncer de mama y lo mismo para los foros de discusión en inglés). A continuación se describen los resultados más significativos (i.e.,  $p < .01$  para evitar el error Tipo 1) para cada una de las categorías analizadas.

Enfoque en sí mismos. Los resultados mostraron que las mujeres de habla hispana y habla inglesa en foros de discusión de depresión utilizan más palabras en primera persona del singular que las mujeres de habla hispana y habla inglesa en foros de discusión de cáncer de mama. Estos resultados replican lo encontrado en otras investigaciones, donde poetas que se suicidaron utilizaban más pronombres en primera persona (Stirman & Pennebaker, 2001), así como lo encontrado por Rude et al. (2004) donde estudiantes deprimidos tendieron a usar más palabras en primera persona que los estudiantes no deprimidos. Estos hallazgos apoyan la teoría de Pyszczynski y Greenberg (1987) que asocia el centrarse en uno mismo con la depresión.

Valencia positiva y negativa. Los resultados mostraron que las mujeres en foros de depresión utilizan más palabras con valencia negativa que las mujeres en foros de cáncer de mama. Solamente en los foros de discusión en inglés se encontró diferencias en

el uso de palabras con valencia positiva. Específicamente las mujeres de foros de cáncer de mama, utilizan más palabras positivas que las mujeres en foros de depresión. Estos resultados están en línea con los hallazgos de otras investigaciones con respecto al uso de palabras negativas y positivas (e. g., Rude et al., 2004). Especialmente en los foros de discusión en inglés. Específicamente, Rude y colaboradores (2004) encontraron que las personas deprimidas tienden a usar más palabras con valencia negativa y menos palabras con valencia positiva que las no deprimidas. Asimismo, los resultados apoyan la teoría de Beck (1967) donde establece que las personas deprimidas tienden enfocarse en aspectos negativos de su vida.

Procesos Sociales. Los resultados mostraron que las mujeres de habla hispana y habla inglesa en foros de discusión de depresión utilizan más palabras en primera persona del plural que las mujeres de habla hispana y habla inglesa en foros de discusión de cáncer de mama. Solamente en los foros de discusión en español se encontró diferencias en el uso de palabras sociales. Específicamente las mujeres de foros de cáncer de mama, utilizan más palabras sociales que las mujeres en foros de depresión. Estos hallazgos apoyan el modelo de suicidio de Durkheim (1951) de que las personas deprimidas tienden a aislarse y a desconectarse del mundo. Aún más, esto va en línea con lo encontrado por Stirman y Pennebaker (2001) de que los poetas que se suicidaron utilizaron menos palabras en primera persona en plural que los poetas que no se suicidaron.

En el Estudio 2 investigamos el lenguaje en foros de discusión en inglés y español. Específicamente, comparamos en ambos foros el lenguaje de mujeres que participaban en foros de depresión y mujeres que participaban en foros de discusión de

cáncer de mama. En general los resultados replicaron lo encontrado en investigaciones previas. Asimismo, el estudio mostró que el uso del lenguaje en mujeres de foros de depresión vs. foros de cáncer de mama es similar en individuos de habla inglesa e hispana. Este hallazgo muestra que la experiencia de depresión es similar en las culturas de habla hispana a las culturas de habla inglesa en la mayoría de las categorías.

### *Limitaciones*

Una limitación de este estudio es la manera en que se definió a las personas deprimidas. Sin duda, no podemos saber a ciencia cierta si las mujeres en los foros de depresión estaban deprimidas y si las mujeres en los foros de cáncer, estaban realmente enfermas de cáncer de mama. Asimismo, no podemos saber si las mujeres en los foros de depresión estaban deprimidas porque tenían cáncer o que en los foros de cáncer las mujeres estaban deprimidas porque tenían cáncer de mama. Sin embargo, se debe tomar en cuenta que aunque la población de este estudio fue diferente a la de Stirman y Pennebaker (2001) y a la de Rude et al. (2004), los resultados fueron similares, lo que da una evidencia más a lo establecido por otros investigadores y por teóricos de la depresión. Otra limitación de este estudio es que no se conocen los datos específicos de los participantes por lo que no se pueden generalizar a otras poblaciones de mujeres.

Futuras investigaciones deben de replicar lo encontrado en esta investigación en una muestra de deprimidos clínicamente y no deprimidos y se deben replicar los resultados en una muestra de hombres.

### **Discusión General**

En esta investigación se han expuesto dos estudios. En el Estudio 1 mostramos el desarrollo del LIWC en español y su equivalencia con el LIWC en inglés. En el Estudio

2, expusimos cómo puede ser utilizado el programa para analizar textos en inglés y español en personas que participan en foros de discusión en Internet. En pocas palabras, introducimos un programa que analiza textos en español, y presentamos su aplicabilidad en el área de la Psicología.

El hecho de que en este estudio se haya encontrado que las mujeres en foros de depresión y mujeres en foros de cáncer de mama utilizan palabras de manera similar en foros de habla inglesa y de habla hispana, demuestra que el LIWC en español puede ser utilizado de la misma manera en que se ha hecho en escenarios de habla inglesa. Por ejemplo, con el LIWC en español se puede estudiar la consistencia del uso del lenguaje a través del tiempo y contextos (e. g., Mehl & Pennebaker, 2003; Pennebaker & King, 1999). Asimismo, se puede emplear para estudiar fenómenos psicológicos tales como la depresión (e. g., Stirman & Pennebaker, 2001; Rude et al., 2004), la personalidad (e. g., Pennebaker & King, 1999) y los aspectos sociales y políticos (e. g., Cohn et al., 2004; Stone & Pennebaker, 2002).

Es importante recordar que el LIWC es fácil de usar y además es muy amistoso para aplicar en distintos contextos. Por ejemplo, para analizar el uso del lenguaje en diversos foros de ayuda en Internet (e. g., Cohn et al., 2004; Stone & Pennebaker, 2002), así como fuentes de texto, literatura, narraciones personales (Pennebaker & Graybeal, 2001) y conversaciones cotidianas (Mehl & Pennebaker, 2003). Más aún, su fácil manejo da la posibilidad de ser adaptado para analizar las idiosincrasias de cada lugar donde se aplique el programa (e. g., el argot cultural), o traumas culturales (e. g., desastres naturales, terrorismo, el asesinato de un líder importante).

Es importante hacer notar que el propósito principal de este manuscrito fue introducir el uso de un programa de computadora que analiza textos en español. De esta manera se puede utilizar este programa para hacer investigación en el área de la Psicología en las culturas de habla hispana. Las investigaciones que se hagan en el futuro no solamente pueden replicar lo hecho en Estados Unidos, sino que también pueden aportar descubrimientos idiosincrásicos de cada lugar donde se hable español. Por ejemplo, en México se podría hacer comparaciones entre distintas clases sociales. Se podría relacionar el uso de las palabras con distintos valores culturales y emocionales. Asimismo, se podría utilizar para observar cuestiones en el área de la salud y otras variables psicológicas importantes. Es realmente imperativo que psicólogos, sociólogos, antropólogos y lingüistas tengan acceso a este programa para que, con sus investigaciones, aporten un mayor entendimiento de los procesos sociales y psicológicos de la cultura hispano-parlante.

De igual manera, se debe resaltar que el LIWC en Español esta en sus inicios, así es que también necesita desarrollarse. Si este programa llega a manos de otras personas que puedan agregar categorías relevantes para cada cultura y además una mejoría a las categorías existentes, se podría crear un diccionario más completo. En pocas palabras ahora se puede empezar a explorar el mundo del uso de las palabras dentro de la apasionante lengua española.

### **Referencias Bibliográficas**

Alpers, G. W., Winzelber, A. J., Classen, C., Roberts, H., Dev, P., Koopman, C., & Taylor, C. B. (2005). Evaluation of computerized text analysis in an Internet breast cancer support group. *Computers in Human Behavior, 21*, 361-376.

- Beck, A. T. (1967). *Depression: Clinical, experimental, and theoretical aspects*. New York: Harper & Row.
- Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Manual for the Beck Depression Inventory-II*. San Antonio, TX: Psychological Corporation.
- Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry, 4*, 561-571.
- Benet-Martínez, V., & John, O. P. (1998). *Los Cinco Grades* across cultures and ethnic groups: Multitrait-multimethod analyses of the Big Five in Spanish and English. *Journal of Personality and Social Psychology, 75*, 729-750.
- Brewer, M. B., & Gardner, W. (1996). Who is the we? Levels of collective identity and self representation. *Journal of Personality and Social Psychology, 71*, 83-93.
- Cohn, M. A., Mehl, M. R., & Pennebaker, J. W. (2004). Linguistic indicators of psychological change after September 11, 2001. *Psychological Science, 15*, 687-693.
- Davison, K. P., Pennebaker, J. W., & Dickerson, S. S. (2000). Who talks? The social psychology of illness support groups. *American Psychologist, 55*, 205-217.
- Durkheim, E. (1951). *Suicide*. New York: Free Press.
- Freud, S. (1901). *Psychopathology of everyday life*. New York: Basic Books.
- Gardner, W. L., Gabriel, S., & Lee, A. Y. (1999). I value freedom, but we value relationships: Self-construal priming mirrors cultural differences in judgment. *Psychological Science, 10*, 321-326.
- Hart, R. P. (1984). *Verbal style and the presidency: A computer-based analysis*. New York: Academic Press.

Hofstede, G. (1983). Dimensions of national cultures in fifty countries and three regions.

En J. B. Deregowski, S. Dziurawiec, & R. C. Annis (Eds.), *Explications in cross-cultural psychology* (pp. 335-355). Lisse: Swets & Zeitlinger.

Joiner, J. T. E., Sandin, B., Chorot, P., Lostao, L., Marquina, G. (1997). Development and factor analytic validation of the SPANAS among women in Spain: (More) Cross-cultural convergences in the structure of mood. *Journal of Personality Assessment, 68*, 600-616.

Markus, H. R., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review, 98*, 224-253.

Mehl, M. R., & Pennebaker, J. W. (2003). The sounds of social life: A psychometric analysis of students' daily social environments and natural conversations. *Journal of Personality and Social Psychology, 84*, 857-870.

Newman, M. L., Pennebaker, J. W., Berry, D. S., & Richards, J. M. (2003). Lying words: Predicting deception from linguistic styles. *Personality and Social Psychology Bulletin, 29*, 665-675.

Páez, D., Martínez, F., Velasco, C., Mayordomo, S., Fernández, I., & Blanco, A. (1999). Validez psicométrica de la escala de alexitimia de Toronto. Un estudio Transcultural. *Boletín de Psicología, 63*, 55-76.

Páez, D., Velasco, C., & González, J. L. (1999). Expressive writing and the role of alexythimia as a dispositional deficit in self-disclosure and psychological health. *Journal of Personality and Social Psychology, 77*, 630-641.

- Pennebaker, J. W., & Beall, S. (1986). Confronting a traumatic event: Toward an understanding of inhibition and disease. *Journal of Abnormal Psychology, 95*, 274-281.
- Pennebaker, J. W., & Francis, M. E. (1996). Cognitive, emotional, and language processes in disclosure. *Cognition and Emotion, 10*, 601-626.
- Pennebaker, J. W., Francis, M. E., & Booth, R. J. (2001) Linguistic Inquiry and Word Count (LIWC) [Software de cómputo].
- Pennebaker, J. W., & Graybeal, A. (2001). Patterns of natural language use: Disclosure, personality, and social integration. *Current Directions in Psychological Science, 10*, 90-93.
- Pennebaker, J. W., & King, L. A. (1999). Linguistic styles: Language use as an individual difference. *Journal of Personality and Social Psychology, 77*, 1296 – 1312.
- Pennebaker, J. W., Mehl, M. R., & Niederhoffer, K. G. (2003). Psychological aspects of natural language use: Our words, our selves. *Annual Review of Psychology, 54*, 547-577.
- Pennebaker, J. W., & Stone, L. D. (2003). Words of wisdom: Language use over the lifespan. *Journal of Personality and Social Psychology, 85*, 291-301.
- Pyszczynski, T., & Greenberg, J. (1987). Self-regulatory preservation and the depressive self-focusing style: A self-awareness theory of depression. *Psychological Bulletin, 102*, 122-138.
- Rude, S. S., Gortner, E. M., & Pennebaker, J. W. (2004). Language use of depressed and depression-vulnerable college students. *Cognition and Emotion, 18*, 1121-1133.

Shirae, E. B., & Levy, D. A. (2004). *Cross-cultural psychology: Critical thinking and contemporary applications*. Boston: Pearson Education Inc.

Stirman, S. W., & Pennebaker, J. W. (2001). Word use in the poetry of suicidal and nonsuicidal poets. *Psychosomatic Medicine*, *63*, 517-522.

Stone, L. D., & Pennebaker, J. W. (2002). Trauma in real time: Talking and avoiding online conversations about the death of Princess Diana. *Basic and Applied Social Psychology*, *24*, 172-182.

Stone, P. J., Dunphy, D. C., Smith, M. S., & Ogilvie, D. M. (1966). *The General Inquirer: A computer approach to content analysis*. Cambridge, MA: M.I.T. Press.

Notas de Pie

<sup>1</sup> Para conocer las palabras que se incluyen en cada categoría del LIWC en inglés, por favor contactar a James Pennebaker ([pennebaker@mail.utexas.edu](mailto:pennebaker@mail.utexas.edu)). Para conocer las palabras que se incluyen en cada categoría del LIWC en español, por favor contactar a Nairán Ramírez-Esparza ([nairan@gmail.com](mailto:nairan@gmail.com)).

<sup>2</sup> Cuando se menciona la palabra de habla hispana o hispano-parlante se refiere a personas que hablan español. Al usar este término no se infiere de dónde son las personas. Pueden ser de cualquier país de Latinoamérica o de España, o incluso hispanos viviendo en Estados Unidos.

Tabla 1. Ejemplo de Cómo Crear un Diccionario en el LIWC

---

%	
1	yo
2	mi
3	mio
4	nosotros
6	nuestro
7	singular
8	plural
9	poseción
10	Refyo
%	
yo	1 10
mi	2 7 10
mí*	3 7 9 10
nuestr*	6 8 9 10
nos	4 8 10
nosotr*	4 8 10

---

Tabla 2. Equivalencia de la Versión del LIWC en Inglés y Español en cada una de las Categorías: Correlaciones, Media, Desviación Estándar y Pruebas t pareadas, y ejemplos por cada categorías

Iniciales de Categorías/ Descripción de Categorías	LIWC (N = 83)									
	r	Español		Inglés		Efecto		Ejemplos de cada categoría		
		Media	DS	Media	DS	t	P			
<b>Dimensiones Lingüísticas Estándares</b>										
CP/Conteo de Palabras	.98	661.33	480.70	642.4	487.65	-1.81	.07			
PPO/Palabras por oración	.89	33.03	49.39	33.03	48.05	-.01	1.00			
SIGNOS?/Oraciones que terminan en ?	.70	.06	.20	0.04	0.13	-1.15	.25			
UNICAS/ Palabras únicas	.92	49.35	10.82	50.3	11.21	2.00	.05			
DIC/ % de palabras capturadas	.87	66.32	7.91	61.85	10.84	-7.23	.00			
SEISLTR/ % de palabras > seis letras	.86	30.05	7.08	25.84	8.32	-8.97	.00			
PRONOM/ Total de pronombres	.85	5.50	3.42	5.68	4.75	.63	.53			Yo, nosotros, tu
YO/ Primera persona singular	.95	.58	1.90	0.91	2.57	3.06	.00			Yo, mío
NOSOTROS/ Primera persona plural	.93	.53	1.01	0.88	1.37	5.43	.00			Nosotros, nuestro
UNOMISMO/ Total primera persona	.94	1.10	2.11	1.78	2.84	5.60	.00			Yo, nosotros, mío
TU/ Total segunda persona	.54	1.54	1.33	1.19	2.3	-1.67	.10			Tu, ustedes
OTRO/ Total tercera persona	.52	2.70	1.56	1.89	2.45	-3.52	.00			Ella, él, ellos
NEGACIÓN/Negaciones	.82	.63	.82	0.65	0.87	.44	.66			No, nunca
AFIRMA/Afirmaciones	.28	.15	.28	0.08	0.18	-2.12	.04			Sí, claro
ARTÍCULO/Artículos	.83	12.33	3.38	8.63	3.1	-17.62	.00			El, la, los, las
PREPO/Preposiciones	.43	16.79	3.18	14.7	2.9	-5.86	.00			A, ante, bajo
NÚMERO/Números	.82	.74	.62	1.07	0.87	6.02	.00			Uno, dos, tres
<b>Procesos Psicológicos</b>										
<b>APECTIVO/Procesos afectivos</b>	.61	4.01	2.53	3.35	1.85	-2.96	.00			Feliz, feo, amargado
EMOPOS/Emociones positivas	.57	2.80	1.81	2.31	1.35	-2.90	.01			Feliz, bonito, bueno
SENTPOS/Sentimientos positivos	.37	.63	1.26	0.33	0.55	-2.31	.02			Feliz, felicidad, amor
OPTIME/Optimismo y Energía	.45	1.01	.79	0.83	0.77	-2.12	.04			Certitud, orgullo, ganar
EMONEG/Emociones negativas	.72	.92	1.19	1.05	1.27	1.21	.23			Odio, enemigo, feo
ANSIEDAD/Ansiedad o Miedo	.75	.18	.32	0.14	0.3	-1.34	.18			Nervioso, miedo, tenso
ENOJO/Enojo	.61	.24	.49	0.35	0.78	1.66	.10			Odiar, matar, enojo
TRISTEZA/Tristeza o depresión	.73	.30	.72	0.33	0.55	.42	.68			Luto, llorar, tristeza

Tabla 2. Continúa

Iniciales de Categorías/ Descripción de Categorías	LIWC (N = 83)											
	Español					Inglés					Ejemplos de cada categoría	
	r	Media	DS	t	P	r	Media	DS	t	P		
<b>MECCOG/Procesos cognitivos</b>												
CAUSA/Causa y Efecto	.67	5.48	2.68	-6.44	.00	3.87	2.9	2.9	-6.44	.00	Causa, saber, debería	
INSIGHT/Entendimiento	.65	.56	.46	2.16	.03	0.65	0.51	0.51	2.16	.03	Porque, efecto, por	
DISCREP/Discrepancias	.50	1.25	.82	-.44	.66	1.2	1.21	1.21	-.44	.66	Pensar, saber, considerar	
INHIB/Inhibiciones	.23	.13	.23	5.39	.00	0.89	1.33	1.33	5.39	.00	Debería, podría	
TENTAT/Tentativos	.54	.23	.37	3.71	.00	0.49	0.75	0.75	3.71	.00	Bloquear, obligar, forzar	
CERTEZA/Certeza	.25	.88	.66	2.14	.04	1.13	0.99	0.99	2.14	.04	Talvez, creo, supongo	
<b>SENTIDOS/Procesos Sensoriales y Preceptuales</b>	.33	.55	.55	2.25	.03	0.72	0.61	0.61	2.25	.03	Siempre, nunca	
<b>SOCIAL/Procesos Sociales</b>												
COMU/Comunicación	.73	1.65	1.01	-4.93	.00	1.22	1.13	1.13	-4.93	.00	Ver, tocar, escuchar	
REFOTRO/Referencia a otras personas	.44	.39	.44	.92	.36	0.44	0.5	0.5	.92	.36	Ver, vista, mirada	
AMIGOS/Amigos	.71	.45	.53	.95	.34	0.51	0.81	0.81	.95	.34	Oído, sonido, escuchar	
FAMILIA/Familia	.55	.22	.28	-.38	.71	0.21	0.26	0.26	-.38	.71	Tocar, sostener, sentir	
HUMANOS/Humanos	.89	7.54	3.85	-2.69	.01	6.88	4.82	4.82	-2.69	.01	Hablar, nosotros, amigos	
<b>Relatividad</b>												
TIEMPO/Tiempo	.75	.96	.75	2.05	.04	1.12	1.08	1.08	2.05	.04	Hablar, compartir, conversar	
PASADO/Tiempo pasado del verbo	.75	4.81	2.51	-2.80	.01	4.06	3.65	3.65	-2.80	.01	1era plural, 2nda, 3era persona	
PRESENTE/Tiempo presente del verbo	.68	.26	.44	-4.03	.00	0.12	0.32	0.32	-4.03	.00	Cuate, amigo, colega	
FUTURO/Tiempo futuro del verbo	.91	.26	.58	-2.04	.04	0.21	0.41	0.41	-2.04	.04	Mama, hermano, primo	
<b>SPACE/Espacio</b>	.85	.97	.97	-3.50	.00	0.75	1.06	1.06	-3.50	.00	Niño, mujer. Grupo	
ARRIBA/Arriba	.71	2.72	1.38	3.66	.00	3.21	1.70	1.70	3.66	.00	Hora, día, noche	
ABAJO/Abajo	.89	1.73	1.63	4.56	.00	2.36	2.49	2.49	4.56	.00	Caminé, fue, tuve	
INCL/Inclusivos	.86	4.45	2.47	3.77	.00	5.35	3.94	3.94	3.77	.00	Camino, es, tengo	
EXCL/Exclusiones	.48	.37	.60	6.37	.00	.94	.91	.91	6.37	.00	Será, haría	
MOCIÓN/moción, movimiento	.47	1.64	1.13	5.12	.00	2.39	1.43	1.43	5.12	.00	Alrededor, arriba, abajo	
	.30	.09	.20	13.51	.00	1.10	.71	.71	13.51	.00	Arriba, encima, alto	
	.05	.09	.20	2.61	.01	.21	.34	.34	2.61	.01	Debajo, abajo	
	.51	5.09	1.35	12.94	.00	7.54	2.00	2.00	12.94	.00	Con, y, incluyendo	
	.36	.68	.69	11.20	.00	2.13	1.24	1.24	11.20	.00	Pero, sin, excepto	
	.32	1.21	.90	-2.797	.00	.91	.77	.77	-2.797	.00		

Tabla 2. Continúa

Iniciales de Categorías/ Descripción de Categorías	LIWC (N = 83)									
	r	Español		Inglés		Efecto		Ejemplos de cada categoría		
		Media	DS	Media	DS	t	p			
<b>Asuntos Personales</b>										
<b>OCUPA/Ocupación</b>										
ESCUELA/Escuela	.86	3.16	2.41	2.97	2.55	-1.27	.21	Trabajar, clase, jefe		
TRABAJO/Trabajo	.81	1.04	1.23	.81	1.41	-2.49	.02	Clase, estudiante, colegio		
LOGRO/Logro	.83	1.53	1.43	1.35	1.41	-1.98	.05	Empleado, jefe, carrera		
<b>LEISURE/Pasatiempos</b>	.56	1.01	.81	1.00	.73	-1.15	.88	Intentar, ganar, objetivo		
CASA/Hogar	.70	.99	1.04	1.08	1.22	.93	.35	TV, música, películas		
DEPORTES/Deportes	.58	.37	.45	.34	.44	-.88	.38	Casa, cocina, refrigerador		
TV/Televisión y cines	.65	.22	.55	.16	.41	-1.35	.18	Fútbol, juego, jugar		
MÚSICA/Música	.82	.20	.36	.17	.38	-1.27	.21	TV, telenovela, programa		
DINERO/Dinero y asuntos financieros	.81	.35	.76	.44	.85	1.60	.11	Tonadas, canciones, cd		
METAFI/Asuntos Metafísicos	.60	.57	.90	.53	.92	-.45	.66	Cambio, dinero, ganancia		
RELIG/Religión	.88	.63	1.00	.65	1.06	.32	.75	Dios, cielo, ataúd		
MUERTE/Muerte	.89	.45	.78	.42	.84	-.71	.48	Dios, cielo, iglesia		
<b>FÍSICO/Estados físicos y funciones</b>	.78	.18	.48	.23	.57	1.08	.28	Muerte, entierro, ataúd		
CUERO/Estados del cuerpo y síntomas	.67	1.95	1.85	1.23	1.75	-4.53	.00	Dolor, pecho, dormir		
SEXUAL/Sexo y sexualidad	.73	.98	1.24	.83	1.47	-1.34	.19	Dolor, corazón, toser		
COMER/Comer, tomar, dieta	.33	.19	.62	.23	.49	.57	.57	Lujuria, pene, sexo		
DORMIR/Dormir, soñar	.14	.59	.42	.13	.29	-8.73	.00	Comer, tragar, probar		
ASEARSE/Acicalarse	.76	.07	.26	.09	.28	.98	.33	Dormitar, cama, sueño		
<b>Otro Grupo de Palabras</b>	.51	.04	.20	.08	.26	1.25	.22	Lavar, baño, limpiar, ducha		
GROSER/Malas palabras, groserías	.54	.02	.07	.00	.03	-2.29	.03	Tonto, cabrón		
RELLENO/ Palabras de relleno	NA	NA	NA	NA	NA	NA	NA	Osearr, esterr, eh,		

Nota. Las palabras en MAYÚSCULA indican las abreviaturas de las categorías como se presentan en el programa en español. Una explicación le sigue a las iniciales. Las palabras en **Negrita** y Subrayadas son las categorías principales (e.g., **Procesos Psicológicos**); las palabras en **Negrita** indican categorías que engloban un grupo de subcategorías (e.g., **AFFECTO** engloba también la subcategoría de Emociones Positivas); las palabras en *Itálica* indican subcategorías de subcategorías (e.g., *Optimismo* es parte de la subcategoría Emociones Positivas), NA significa que no se aplica para este estudio; todas las correlaciones son significativas al .01, excepto por aquellas subrayadas ( $p = 0 > .05$ )

Tabla 3. Diferencias en el Uso de Palabras en Foros de Discusión en Inglés y en Español entre el Grupo Deprimido y el Grupo Control.

Categorías	Foros de Discusión				Efectos Principales	
	Español		Inglés		Español	Inglés
	<u>Depresión</u>	<u>Cáncer</u> <u>Mama</u>	<u>Depresión</u>	<u>Cáncer</u> <u>Mama</u>		
	<u>N = 80</u> <u>M(DS)</u>	<u>N = 80</u> <u>M(DS)</u>	<u>N = 80</u> <u>M(DS)</u>	<u>N = 80</u> <u>M(DS)</u>	<i>t</i>	<i>t</i>
<b>Enfoque en sí mismos</b>						
Yo/ Primera persona singular	9.30 (2.34)	5.03 (2.76)	12.24 (2.97)	4.03 (3.01)	-10.54*	-17.39*
<b>Valencia positiva/negativa</b>						
Emociones positivas	2.99 (1.36)	3.53 (1.93)	1.72 (1.14)	2.54 (1.72)	2.04 <sup>+</sup>	3.56*
Emociones negativas	3.64 (1.56)	1.93 (1.23)	3.34 (1.48)	1.18 (1.06)	-7.69*	-10.59*
<b>Procesos Sociales</b>						
Primera persona plural	.18 (.33)	.72 (1.06)	.22 (.39)	1.02 (1.28)	4.36*	5.32*
Palabras Sociales	9.3 (2.37)	11.53 (3.63)	7.03 (4.24)	7.87 (4.12)	4.56*	1.27

Nota: Se presentan medias, desviaciones estándar, efectos principales para los foros de discusión en inglés y los foros de discusión en español; las pruebas t-independiente son de dos colas y se basan en 158 grados de libertad; <sup>+</sup> p < .05, \* p < .001.