



UNIVERSIDAD DE BELGRANO

Las tesis de Belgrano

**Facultad de Humanidades
Carrera de Licenciatura en Psicología**

**Técnicas de análisis de textos asistidas por
computadoras para investigaciones en
psicología clínica: Diseño de un diccionario
ponderado en español para la Actividad
Referencial**

Nº 164

Manuela O'Connell

Tutor: Andrés Roussos

Departamento de Investigación
Junio 2005

Agradecimientos

Quiero agradecer enormemente a Daniel quien me apoyó, me escuchó y me ayudó en todo este proceso de hacer y escribir la tesina. Pero fundamentalmente quiero agradecerle por haberme acompañado en esta difícil pero muy feliz decisión de haberme adentrado en el oficio de ser terapeuta.

A mi tutor, Andrés Roussos, quien me guió y enseñó a lo largo de todo este proceso de trabajo. Su excelente dirección hizo posible la finalización de esta tarea.

A los miembros del equipo de investigación dirigidos por Andrés Roussos.

A Demián quien desde la panza me acompañó inclusive en clases y exámenes.

A mis padres por su amor incondicional.

Y finalmente, como siempre, a Luna (ya vas por la segunda, espero que no haya una tercera...).

Índice

RESUMEN	7
INTRODUCCIÓN	7
Objetivos del trabajo y justificación de la temática	7
Análisis del contenido de textos asistido por computadoras	8
Hipótesis de trabajo	9
Pilares de las técnicas de análisis de textos asistidas por computadoras	9
Diccionarios	10
Constructos psicológicos evaluados con diccionarios a través del uso de técnicas asistidas por computadoras	11
Actividad Referencial	12
Medidas de AR	13
MÉTODOS	15
Muestra	15
Materiales	17
Procedimiento	17
RESULTADOS	22
Descripción de WRAD en castellano	22
Comparación con el diccionario en inglés	22
Análisis del diccionario	22
Categorías de RA	22
Correlación interna	24
Correlación de prueba	25
Correlación externa	26
Cobertura	26
Comparación con los resultados obtenidos para WRAD	26
DISCUSIÓN Y CONCLUSIONES	27
REFERENCIAS	32
ÍNDICE DE TABLAS Y FIGURAS	
Tabla 1.- Esquema de las características de las sesiones utilizadas	16
Tabla 2.- Síntesis del material utilizado	17
Tabla 3.- Categorías de la AR utilizadas	18
Tabla 4.- Las 20 primeras filas de la matriz principal	19
Tabla 5.- Ponderación para cada categoría dominante absoluta	20
Tabla 6.- Asignación de la ponderación en el caso de dos categorías dominantes.	20
Tabla 7.- Las 20 primeras palabras más frecuentes en castellano y en inglés.	22
Figura 1.- Cantidad de palabras en cada categoría de RA.	23
Figura 2.- Cantidad de veces que los jueces calificaron con los distintos puntajes	23
Figura 3.- Cantidad de palabras de la unidad versus calificación de los jueces en escala decimal y en escala logarítmica.	24
Figura 4.- Correlación interna para los distintos P con la suma.	24
Figura 5.- Correlación interna para distintos P para el promedio MSWRAD.	25
Figura 6.- Correlación de prueba para la selección del parámetro de ponderación P.	25
Figura 7.- Correlación para la selección de P sobre los archivos S1 y S6.	26
Tabla 8.- Comparación de los resultados obtenidos en inglés y en castellano.	26
Figura 8.- Gráficos de estructura para el promedio con el conjunto de textos utilizados para la construcción del diccionario para distintos parámetros de ponderación.	28
Figura 9.- Gráfico de la estructura para el conjunto de textos de selección del parámetro de ponderación.	29
Figura 10.- Gráfico de la estructura para el conjunto de textos no utilizados en la construcción del diccionario.	29
Figura 11.- Gráfico de puntos para la calificación de los jueces versus diccionario. Cada texto distinto se incluye en colores distintos.	30

Resumen

El objetivo de este trabajo es presentar la construcción de una versión en español del Diccionario de Actividad Referencial ponderado. Este instrumento, desarrollado por Bucci y Maskit, es utilizado para modelar computacionalmente una variable psicolingüística: la Actividad Referencial (AR) en lengua hablada. La dimensión de la Actividad Referencial concierne al grado con el que el lenguaje refleja conexión con la experiencia no verbal, incluyendo la presentación en imágenes y la experiencia corporal y emocional, y evoca una experiencia acorde en el que escucha o el que lee. La AR está primariamente indicada por los atributos de estilo del lenguaje independiente de los de contenido. Una AR alta indica un lenguaje vívido y evocativo, y una AR baja indica un lenguaje general, abstracto, vago o difuso. Las escalas de AR han sido ampliamente utilizadas en investigaciones psicolingüísticas y clínicas.

La medición del proceso referencial utilizando el S-WRAD provee una base para el análisis estructural de la dinámica de las intervenciones en relación a las reacciones lingüísticas del paciente.

En la presente tesina, se obtuvieron resultados estadísticos similares a la versión inglesa. El diccionario de Actividad Referencial ponderado que se generó consta de 470 ítems léxicos, con ponderaciones con valores que oscilan entre -1 y 1 . Para su elaboración, se realizaron los pasos detallados para el armado del diccionario establecidos por Bucci y Maskit. En el presente trabajo se discuten también aspectos técnicos sobre la metodología estadística y general utilizada en la construcción del mismo¹.

Introducción

Objetivos del trabajo y justificación de la temática

El objetivo general de este trabajo es presentar y estudiar el desarrollo de técnicas de análisis de textos asistidas por computadoras. Los objetivos específicos llevados a cabo han sido:

- a) Construir un diccionario ponderado para la Actividad Referencial en español (S-WRAD) en base a la réplica de Bucci y Maskit (2004).
- b) Sistematizar los procedimientos para analizar acuerdos entre dos modalidades de medición de constructos psicológicos para su análisis a través de técnicas asistidas por computadoras.
- c) Evaluar los procedimientos para la modelización de distintos estilos lingüísticos a través del uso de diccionarios.

La investigación en psicología clínica se vale de múltiples herramientas con el fin, entre otras cosas, de estudiar y analizar el proceso terapéutico. Estas herramientas pueden generarse a partir de distintas modalidades de análisis, una de ellas, que es la que se utiliza en este trabajo, es el análisis de la narrativa. Dentro de las distintas modalidades de análisis de la narrativa, una forma posible de estudiar los procesos terapéuticos y sus elementos concomitantes como intervenciones del terapeuta, proceso del paciente, etc; es a través del análisis del lenguaje del paciente (entendiendo a esta como un análisis acotado del fenómeno narrativo), tal como afirma Rosow (1997). Se pueden abrir nuevas posibilidades de entendimiento y de preguntas si se enfoca en las propiedades del lenguaje y en el modo en que las palabras representan procesos mentales.

Dentro de los posibles modos de análisis del lenguaje, se pueden diferenciar los métodos cualitativos y cuantitativos. En esta tesina se presenta una modalidad de análisis cuantitativa, como es el análisis de texto asistido por computadoras, sosteniendo que este tipo de análisis es una de las herramientas que posibilita el estudio de los procesos terapéuticos de manera objetiva y sistemática.

Existen actualmente diversas técnicas cuantitativas de análisis del lenguaje. En particular en este trabajo, se estudian aquellas que son asistidas por computadoras. Dentro de ellas, algunas utilizan diccionarios técnicos para el estudio de ciertas variables psicológicas que puedan dar información acerca del proceso terapéutico. Los procedimientos computacionales son apropiados para analizar muestras grandes así como estudios longitudinales, a la vez que permiten un seguimiento micro analítico confiable de las fluctuaciones de una variable psicológica en diversas formas de discurso, tal como Bucci y Maskit (2004) establecen para la Actividad Referencial.

Hasta el momento los diccionarios desarrollados, es decir construidos y probados, en castellano para el análisis de texto asistido por computadoras no son ponderados; esto quiere decir que un ítem se encuentra o no en el diccionario. Los diccionarios ponderados son particularmente importantes para analizar las variables de estilo que varían en distintos grados en contraste con los elementos de contenido que usualmente se los puede definir como ausentes o presentes. Este tipo de diccionario permitiría un análisis más continuo de la variable psicológica de interés posibilitando un análisis del proceso terapéutico también continuo.

Dentro de las variables psicológicas posibles que se presentarán más adelante se elige para este trabajo la Actividad Referencial. Esta elección se encuentra basada en la importancia de este constructo por su facilidad de aplicación y profundidad de análisis del proceso terapéutico. Bucci y Maskit (2004) desarrollaron un diccionario ponderado para la Actividad Referencial en inglés (WRAD) pero no existe una versión en español del mismo, el cual permitiría la aplicación de esta teoría y el análisis de procesos terapéuticos dentro de esta lengua.

Por lo tanto, el horizonte final de los objetivos establecidos en esta tesina es su posterior aplicación en investigaciones clínicas.

Con relación a los potenciales usos del diccionario, el ámbito principal del mismo es el uso en investigación en psicoterapia. Permite tanto la evaluación del discurso del paciente como del terapeuta, analizando los estilos lingüísticos de los mismos y estableciendo su vínculo con otros estilos lingüísticos o peculiaridades con patologías o modalidades de tratamiento (Sammons & Siegel, 1998 y Roussos y Leibovich de Duarte, 2002).

Análisis del contenido de textos asistido por computadoras

La necesidad de realizar investigaciones que den cuenta de aspectos claves del accionar terapéutico, como pueden ser su eficacia y su pertinencia para las distintas patologías mentales, genera al mismo tiempo la necesidad de diseñar nuevos instrumentos que permitan realizar análisis de estas características. Entre las múltiples técnicas propuestas para este tipo de estudios, se encuentran aquellas que son asistidas por computadoras (Roussos, Acosta, Juárez y Mergenthaler, 2001). En ellas se puede recurrir a la utilización de computadoras ya sea para la obtención, análisis o presentación de los datos.

Este tipo de técnicas se están incorporando a la práctica de los profesionales de la salud mental, tanto para el análisis de aspectos clínicos como para la investigación básica y aplicada en esta área (Roussos, Acosta, Juárez y Mergenthaler, 2001).

La introducción de nuevas técnicas en el ámbito profesional obedece a distintas razones, entre ellas, la necesidad de brindar nuevos tipos de respuesta relativos al accionar psicoterapéutico. Dentro de las posibles técnicas asistidas por computadora se encuentran aquellas que utilizan como objeto de análisis textos clínicos, es decir, realizan un análisis del lenguaje de materiales transcritos de diadas pacientes-terapeutas o entrevistados-entrevistador.

El análisis de textos terapéuticos asistidos por computadora es una de las nuevas herramientas que se han diseñado para profundizar los estudios empíricos en nuestro ámbito de trabajo. Estos intentos pretenden otorgarle al quehacer psicoterapéutico elementos objetivos que den cuenta de la eficacia de su accionar (Roussos, Acosta, Juárez y Mergenthaler, 2001).

El alto grado de innovación que reflejan estas técnicas se basa, por un lado, en que brindan la posibilidad de estudiar en forma objetiva el proceso terapéutico y sus componentes, y por otro, en que permiten la inclusión de nuevas tecnologías para este tipo de investigación, facilitando el análisis de indicadores de eventos psicoterapéuticos muy complejos y que hasta no hace mucho tiempo era imposible considerar (Roussos, Acosta, Juárez y Mergenthaler, 2001).

Hay distintas técnicas de análisis de textos que varían en el grado de sofisticación y que pueden o no ser asistidas por computadoras. El primer método es cualitativo pero sus resultados son cuantificables y consiste en la calificación mediante jueces. En este sistema jueces entrenados dividen el texto en partes y lo clasifican basándose en reglas que tienen en cuenta la sintaxis, el contenido y el estilo (Rosow, 1997).

La segunda técnica es un sistema que cuenta palabras individualmente y que luego compara el texto con una lista de palabras. Cada palabra en la lista tiene una descripción categorial que cuando se la compara con las palabras del texto provee frecuencias y otros procedimientos estadísticos. A este se lo denomina sistema de contabilización de palabras aisladas (Rosow, 1997). Esta técnica para que sea más eficiente puede ser asistida por computadoras como por ejemplo el CRA creado por Mergenthaler y Bucci (1999) que se explicará más adelante.

La tercer técnica computarizada es la "desambiguación", que es un procedimiento más sofisticado que recurre a características de la inteligencia artificial. Este procedimiento es sensible al contexto y va más allá de contar palabras al localizarlas en las estructuras sintácticas. Se emplea para determinar y después especificar el significado de las palabras utilizando algoritmos que se basan en palabras vecinas importantes. Un ejemplo de esta aproximación es el «General Inquirer Program» que se usa junto con el «Harvard IV Psychosocial Dictionary» y el sistema alemán EVA (Grunzig, Holzschek & Kachele, 1976).

En 1990 se realizó un estudio que compara estos tres métodos de análisis de textos aplicándolos a discursos de grupos de pacientes psiquiátricos. El resultado fue que el sistema de conteo de palabras provee una clasificación diagnóstica más precisa, lo cual implica que proporciona una medida confiable en la clasificación de los textos a pesar de la simplicidad del método (Rosow, 1997).

El resultado y las conclusiones de este estudio fueron justificadas por Spence (1980) argumentando que en las conversaciones las personas se enfocan en el significado en vez de en la elección de las palabras. La elección de las palabras depende de un proceso inconsciente que revela potencialmente temas subyacentes como necesidades y motivaciones. Por lo tanto, el conteo de palabras en forma aislada es mejor porque al colocar la atención en el contexto se reduciría la sensibilidad a los temas subyacentes, como lo afirma Spence (citado en Rosow, 1997). Por esta razón, en la presente tesina se han seleccionado dentro de las distintas técnicas de análisis de textos asistidas por computadoras, aquellas que contabilizan palabras.

La utilización de este tipo de análisis de textos asistidos por computadoras descansa en la premisa de que las características emocionales y aspectos lingüísticos significativos se pueden identificar cuantitativamente basándose en ítems léxicos (Rosow, 1997).

En el caso del presente trabajo, el análisis de textos asistidos por computadora que se estudia, es un proceso mediante el cual se genera una lista de palabras que previamente han mostrado ser representativas de ciertos contenidos específicos o estilos del lenguaje. Estas palabras agrupadas en «diccionarios» se comparan con los textos.

Entre una de las técnicas de este tipo ya desarrolladas se puede citar el Modelo de Ciclo Terapéutico desarrollado por Erhard Mergenthaler en la Universidad de Ulm Alemania, que posee la peculiaridad de promover la incorporación de sistemas computarizados para el análisis de contenidos del discurso en el desarrollo de estudios sobre procesos terapéuticos.

Estos procedimientos tienen ventajas sobre la evaluación mediante jueces. Permiten calificar una base de textos más grandes ya que reducen el tiempo y la energía necesaria para realizarlos. Además se elimina el factor de la confiabilidad de la calificación.

Hipótesis de trabajo

Como se mencionó anteriormente, dentro de las técnicas de análisis de estilos lingüísticos de textos existen en particular dos métodos que se discutirán y compararán en este trabajo: la calificación mediante jueces y la medición mediante una computadora. Cuando se trata de asistir mediante computadoras el análisis de estilos lingüísticos, se formula un determinado procedimiento que culmina en la medición de una determinada variable psicológica. Es interesante pensar si la medición por jueces y por computadora de una determinada variable psicológica, «miden» exactamente lo mismo. Como se verá más adelante, para poder discriminar esto, es necesario definir la variable psicológica que se utiliza y la operacionalización de la misma.

Por lo tanto, a la hora de usar técnicas de análisis de textos asistidas por computadoras, para su posterior comparación con otros métodos, es necesario ser riguroso y explícito con la manera en que se toma una determinada variable psicológica.

Por otro lado, es interesante plantear si al momento de desarrollar una de estas técnicas se trata de emular el procesamiento humano realizado por los jueces o en que se basa y justifica su construcción. Se podría estimar, en el caso de técnicas de análisis de estilos lingüísticos de textos de palabras aisladas, que para la calificación los jueces implícitamente analizan ciertas palabras características pero sería un error reducirlo solamente a esto. Es por esto que, con el fin de poder establecer comparaciones, en este trabajo se toma como hipótesis que tanto los jueces como el procedimiento computacional miden lo mismo (este punto será replanteado y discutido en la conclusión).

Pilares de las técnicas de análisis de textos asistidas por computadoras

Las técnicas de análisis de estilos lingüísticos de textos asistidos por computadoras, como la desarrollada en este trabajo, descansa en tres pilares:

- el constructo psicológico a estudiar,
- el diccionario técnico específico,
- el programa o plataforma sobre la que se asienta.

Esto significa que esta técnica implica la definición y el recorte necesario del constructo psicológico para la operacionalización de una cierta variable psicológica. Además, es necesario la conformación y/o utilización de un diccionario específico que mida esta variable (por otro lado, hay distintos tipos posibles de diccionarios). Para la aplicación de dicho diccionario, con el fin de medir la variable a estudiar, se necesita un programa que lo utilice y lo «aplique» al texto de una manera particular. Si bien todos los programas computacionales para analizar textos terapéuticos emplean diccionarios técnicos, dichos diccionarios y programas tienen una independencia entre sí. Esto quiere decir que un mismo diccionario se puede usar sobre diferentes programas o que un mismo programa puede utilizar distintos diccionarios. Un ejemplo de este último caso es de Ciclo Terapéutico antes mencionado. Sin embargo, no se puede utilizar un diccionario sin un programa ni un programa de este tipo sin un diccionario. A su vez, tanto los diccionarios como los programas dependen de la definición y operacionalización de la variable psicológica a estudiar. Distintos

tipos de definiciones de una misma variable pueden generar distintos tipos de diccionarios o de modos de analizarlos. Ejemplos de esto se verán más adelante. En conclusión, los tres componentes fundamentales del tipo de técnica de análisis de texto asistida por computadoras que se desarrolla en este trabajo son interdependientes entre sí, en el sentido en que según el tipo de definición de la variable o tipo de diccionario o programa que se quiera utilizar se debe elegir un tipo compatible de los otros elementos. Esto quiere decir que una vez efectuada la definición de una determinada variable psicológica y su operacionalización ya no se pueden elegir dentro de todos los tipos de diccionarios posibles, sino que se deben sólo tomar aquellos que puedan dar cuenta de la variable así definida. Por otro lado, dependiendo del diccionario que se eligió, se lo debe asentar sobre un determinado programa que tenga en cuenta sus características.

Diccionarios

Los *diccionarios* técnicos son aquellos que conectan el vocabulario de un texto con los intereses teóricos de quien se encuentra realizando una observación. Se definen como una lista de palabras con un sistema categorial. En sí, son el componente formal del modelo de análisis de texto asistido por computadora (Roussos, Acosta, Juárez y Mergenthaler, 2001).

Los principios básicos para la confección de los diccionarios son lo que Mergenthaler denomina *perspectiva de la marca*. Esta se define como un conjunto de indicadores de la presencia de ciertos rasgos lingüísticos (tal como palabras que marcan partes del código del lenguaje o conjuntos de sufijos que indican la presencia de contenidos específicos). Entonces, las palabras que el paciente utiliza en las sesiones se convierten en indicadores de momentos específicos del proceso terapéutico (Mergenthaler, 1996).

Los diccionarios nunca son exhaustivos (o totalmente completos) dado que no se toman todas las expresiones posibles de cada estilo a analizar sino una muestra de las mismas. En el proceso de recorte del lenguaje, que supone la elaboración de un diccionario, existen numerosas perspectivas posibles de selección de palabras (Roussos, Acosta, Juárez y Mergenthaler, 2001).

Un ejemplo de diccionarios lingüísticos son aquellos utilizados por el TAS (Text Analysis System) que es un programa para computadoras creado por Mergenthaler (1992) con el fin de obtener información sobre la frecuencia de determinadas palabras en los textos terapéuticos analizados. Este análisis se realiza desde el marco teórico del *Cycles Models* (CM) que se describirá más adelante. Primero, existió el Modelo de Ciclo Terapéutico que constaba de dos diccionarios (tono emocional y abstracción). Luego a este modelo se le agregó el diccionario de Actividad Referencial constituyendo así el Modelo de Ciclos Terapéuticos. El CM se basa en la detección de tres estilos lingüísticos: el tono emocional, el lenguaje de abstracción y la Actividad Referencial siendo este el primer diccionario desarrollado para tal fin. Este software utiliza tres diccionarios técnicos con base en tales estilos. Estos diccionarios pueden ubicarse dentro de lo que se define como la *perspectiva de la marca en el análisis de contenidos*. Es decir, el interés de estos diccionarios radica en que posibilitan conocer el porcentaje y la distribución de palabras implicadas en un texto particular, sin tomar en cuenta cuál es el contenido específico de las mismas (Roussos, Acosta, Juárez y Mergenthaler, 2001).

El Modelo de Ciclo Terapéutico (MCT) fue desarrollado por Mergenthaler posibilita el estudio del proceso psicoterapéutico a través de la detección de sus momentos y sesiones claves. Estos momentos se definen como situaciones clínicamente importantes, ya que constituyen puntos de quiebre dentro de la sesión o del tratamiento y son necesarios para lograr el cambio en el actuar del paciente (Mergenthaler y Bucci, 1996). Se detectan a través de eventos lingüísticos asociados a los mismos y los eventos se localizan mediante el conteo de palabras pertenecientes a los diccionarios incorporados al modelo (Mergenthaler, 1996).

Para realizar las mediciones de las frecuencias de palabras que permitan identificar elementos del MCT, se utiliza la asistencia de distintos programas de análisis de textos. Los requisitos para dicho análisis se le plantean al sistema a través de la inclusión de parámetros de acción, denominados diccionarios técnicos. En el caso del MCT se reducen las posibilidades a los dos (los dos primeros) o tres (dependiendo si es el Modelo de Ciclo Terapéutico o el de Ciclos Terapéuticos) parámetros fijos de análisis: el tono emocional, el lenguaje de abstracción y la Actividad Referencial perteneciente al modelo de procesos referenciales (Mergenthaler, 1996). Estos parámetros son medidos por el TAS, a través de 3 diccionarios técnicos: el diccionario de Emoción, el diccionario de Abstracción y el diccionario de Actividad Referencial Computarizada (CRA) (Roussos, Acosta, Juárez y Mergenthaler, 2001).

Una vez confeccionados los diccionarios es necesario establecer la forma de introducción del material clínico al sistema. En otras palabras, todos los diccionarios funcionan sobre una plataforma. Por ejemplo los diccionarios del modelo de Ciclo Terapéutico funcionan sobre el TAS que es un programa diseñado para lograr la selección y el conteo de palabras presentes en un texto. Estas palabras se encuentran previamente tipificadas en cada uno de los diccionarios. Una vez realizados los procesamientos el programa muestra un gráfico que representa la distribución de los tres estilos terapéuticos definidos por el MCT (Roussos, Acosta, Juárez y Mergenthaler, 2001).

Constructos psicológicos evaluados con diccionarios a través del uso de técnicas asistidas por computadoras

Todos los programas y diccionarios técnicos evalúan ciertos constructos psicológicos. Para poder realizarlo es necesario definirlos tanto conceptual como operacionalmente. Según como sea esa definición existen diversas estrategias posibles para abordarlo y de la elección de esta estrategia depende cuál diccionario se utiliza y de qué modo.

Para ilustrar esto, se toman distintos constructos que han sido evaluados mediante diccionarios técnicos.

- Tono emocional

Hay distintas maneras de clasificar la emoción. Una de ellas es la propuesta por Dahl, Holzer y Berry (Dahl et al., 1992). Este modo de abordar el constructo de la emoción descansa sobre tres conceptos fundamentales: los deseos, las creencias y las necesidades. El marco teórico de esta variable es la Teoría de Emociones de Dahl.

También se puede considerar a la variable «emoción» como aquella que incluye palabras que revelan un estado emocional o afectivo en el hablante, capaz también de provocar emoción en quien escucha.

Dentro de esta clasificación de la variable emoción, uno de las posibles maneras de evaluación es mediante el MCT. Se atiende únicamente a los indicadores de emoción que pueden detectarse en una transcripción. Por lo tanto, el concepto de emoción se entiende como el tono emocional de un texto. Lo que mide es la densidad con que un texto presenta lenguaje emocional (Roussos, Acosta, Juárez y Mergenthaler, 2001).

Debido a la importancia central de la emoción dentro del proceso psicoterapéutico, se asume que la presencia del tono emocional es un prerrequisito necesario para la aparición de los momentos clave de una sesión o proceso terapéutico.

Otro modo posible de medición de la variable emoción, es a través del análisis de textos asistidos por computadoras, creado por Bucci de manera conceptual. Para ello se recaudó una lista de palabras de distintas fuentes como del «Regressive Imaginary Dictionary» (Martindale, 1975) y del «Affective Lexicon» (Clore et al., 1987). Luego, un grupo de un millón de ítems léxicos fueron analizados con una connotación emocional. De esta manera se creó una lista de palabras inicial. El diccionario fue revisado utilizando diversos criterios de exclusión para finalmente conformar el diccionario final. Uno de estos criterios de exclusión fue el de las palabras concretas y de aquellas con múltiples sentidos (Rosow, 1997).

- Abstracción

La variable abstracción se definió para captar las palabras que indiquen reflexión y pensamientos abstractos. Un alto nivel de abstracción es un indicador de pensamientos complejos y evaluaciones lógicas.

La variable abstracción se entiende como el concepto que lleva al desarrollo del entendimiento y la percepción o como el mecanismo central que conduce a la adquisición de nuevas estructuras. Esta es otra variable que se evalúa en MCT (Roussos, Acosta, Juárez y Mergenthaler, 2001).

Operacionalmente se clasifican a las palabras como abstractas cuando reúnen al menos uno de los siguientes criterios:

- la palabra hace referencia a una generalidad, a un conjunto de entes y no a uno en particular,
- aquello a lo que hace alusión la palabra no es susceptible de ser captado por nuestros sentidos.

Desde un punto de vista lingüístico, la abstracción se define como habilidad de construir términos abstractos a partir de conceptos concretos realizando una operación de transformación morfológica (ej: tierno-ternura).

El diccionario que mide la abstracción (AB) consiste en sustantivos abstractos complejos que son representativos de la reflexión lógica y de la evaluación. Fue confeccionado sólo de manera conceptual y no de manera empírica. Esta lista de palabras se construyó en idioma inglés y alemán mediante un procedimiento automatizado de análisis de sufijos, utilizando el TAS (Roussos, Acosta, Juárez y Mergenthaler, 2001). En idioma español no fue posible repetir estos procedimientos.

- Función reflexiva

El constructo función reflexiva (Fonagy et al., 2002) muestra la capacidad metacognitiva de una persona para mentalizarse acerca de sí mismo en término de motivaciones internas, deseos, afectos e intenciones. Involucra un componente intrapersonal y un componente interpersonal ya que afecta y regula el funcionamiento interpersonal. Se ha desarrollado un programa computacional de análisis de textos por Fertuck et al. que mide este constructo con el propósito de estudiar los casos de personas borderline (Fertuck et al., 2002). Este diccionario se encuentra en elaboración.

Actividad Referencial

La *Actividad Referencial* es un concepto planteado por Bucci (1997) que da cuenta de la capacidad del ser humano de expresar niveles de experiencia no verbal, incluyendo experiencias emocionales, en forma verbal. El hablante o escritor transforma la experiencia interna en una forma verbal; y el oyente o lector transforma las palabras del otro en su propio sistema representacional no verbal (Mergenthaler y Bucci, 1999). En otras palabras, es la actividad de las conexiones referenciales entre los sistemas verbales y no verbales en ambas direcciones (Bucci, 1997)

La AR es un concepto teórico de la *teoría de codificación múltiple*, un modelo del procesamiento de la información emocional derivado del trabajo en psicología cognitiva (Mergenthaler, 1996). La teoría de los «códigos múltiples» se refiere a la interacción existente entre los sistemas emocionales, cognitivos, lingüísticos y orgánicos. Según esta teoría, existen tres formas básicas a través de las cuales el ser humano representa y procesa información (Bucci, 1993). Estas son: el sistema subsimbólico no verbal, el sistema simbólico no verbal y el sistema simbólico verbal. El sistema subsimbólico es el que hace referencia a los procesos que ocurren a nivel orgánico, viscerales y motores, incluye las formas sensoriales de recepción de olores, sonidos, sabores y otras sensaciones. Es un proceso de tipo «analógico» que representa sucesos de orden continuo en lugar de sucesos discretos aislables como podría ser la palabra en el sistema verbal.

El sistema simbólico se caracteriza porque sus componentes, los símbolos, tienen características de ser entidades discretas. La función de los símbolos es la representación de otros entes, además de su propia representación, lo que los vuelve a sí mismos eventos simbolizados. Esto permite que se vuelvan entes generativos, y que se puedan combinar generando una infinita variedad de nuevos eventos (Bucci, 1997).

El sistema simbólico tiene dos niveles:

El simbólico no verbal, representado clásicamente por imágenes, visuales o auditivas, como puede ser la representación mental de una melodía. Estas imágenes tienen la característica de ser unidades discretas, posibles de ser delimitadas y secuenciadas.

El sistema simbólico verbal, que está fundamentalmente centrado en el lenguaje, en el cual las palabras son las unidades que permiten componer, a través de nuevas configuraciones, un esquema de representaciones (Bucci, 1997).

Cada uno de los sistemas tiene sus respectivos contenidos y sus principios organizadores particulares. A su vez estos tres sistemas se encuentran interconectados en el proceso referencial, el cual vincula a los sistemas no verbales entre sí y con el sistema verbal (Bucci, 1997). Es a partir de esto que la autora define la Actividad Referencial (AR).

Esta teoría explica el cambio terapéutico como resultado de una acción sobre los vínculos existentes entre los sistemas subsimbólicos, simbólico no verbal y simbólico verbal (Roussos, Acosta, Juárez y Mergenthaler, 2001).

En general, el proceso de verbalización de las emociones es complejo. El «*Proceso referencial*» es un modelo descrito por Bucci para convertir la información no verbal, incluyendo los esquemas emocionales, en lenguaje. En el mismo, se pueden identificar tres grandes fases en el proceso de exploración de la experiencia emocional y la comunicación de la misma en un contexto interpersonal, como por ejemplo en la psicoterapia. Una *fase inicial* de excitación, en la cual domina la experiencia subsimbólica en formato sensorial, visceral o motor. El individuo puede hablar de sus luchas internas y también de los eventos sensoriales y somáticos que ha experimentado. Un *segundo momento* es aquel en el que emerge en la conciencia una imagen o un episodio que puede ser reciente, un sueño o un recuerdo remoto que represente los contenidos problemáticos y que pueda ser contado en forma narrativa. Una tercera fase de reflexión y explicación. Estas tres fases pueden suceder recurrentemente a lo largo de una sesión o de un tratamiento formando lo que se conoce como el «Proceso Referencial», que permite el acercamiento del individuo a su experiencia emocional posibilitando el cambio de cómo esa experiencia es percibida o entendida (Mergenthaler y Bucci, 1999).

El proceso referencial varía en cualidad y actividad a lo largo de un tratamiento. La Actividad Referencial puede variar por muchas razones. Por ejemplo, dado que la AR es una medida de la traducción de la experiencia interna en palabras, es particularmente sensible a estados internos de ansiedad y depresión y puede ser también afectada por contextos externos como relaciones interpersonales (Rosow, 1997). Las variaciones se muestran en los detalles del estilo del lenguaje. Las conexiones referenciales son más activas para palabras correspondientes a imágenes específicas y concretas, y menos activas para términos y conceptos generales y abstractos (Mergenthaler y Bucci, 1999).

Dado que existen patrones lingüísticos que reflejan directamente la fuerza de las conexiones referenciales, la AR resulta una entidad cuantificable a través de estos patrones que proveen el contexto de los procesos referenciales. En este sentido, se vuelve posible estudiar la psicoterapia más efectivamente utili-

zando patrones lingüísticos en vez de contenidos. Para ello, se desarrollaron escalas de lenguaje específicas, que se describirán luego (Rosow, 1997).

La Actividad Referencial puede ser empleada para el análisis y comprensión del proceso terapéutico, pudiéndose realizar análisis tanto en niveles «micro» como «macro», ya sea que se analicen sesiones aisladas o procesos terapéuticos completos. La Actividad Referencial varía en el transcurso del proceso terapéutico, manteniendo niveles bajos cuando el paciente no ha integrado aún la experiencia en forma lingüística. Los niveles de Actividad Referencial se elevan cuando esta integración se concreta (Roussos, Acosta, Juárez y Mergenthaler, 2001).

Medidas de AR

Las medidas de AR evalúan el grado en que un emisor verbal o escritor es capaz de traducir dicha experiencia en palabras de modo que evoque experiencias correspondientes para el lector o receptor. Utilizando los procedimientos de análisis del estilo del lenguaje, se desarrollaron indicadores observables de los componentes del ciclo referencial. Los trabajos en este área se basan en mediciones lingüísticas no centradas en el contenido del discurso sino en identificar los patrones lingüísticos y las representaciones subjetivas presentes. Estas mediciones se basan en la premisa de redundancia en las expresiones emocionales. Bucci la explica postulando que lo que se revela en el tono de la voz y en las expresiones faciales también tiende a reflejarse en algunas formas del lenguaje (Rosow, 1997).

Para la conceptualización y medición de la Actividad Referencial, Wilma Bucci (1992) estableció cuatro escalas, desarrolladas para evaluar el grado en que la emoción y la fantasía se expresan en palabras: Concretud, Especificidad, Claridad y Presentación de Imágenes. Las escalas reflejan estándares derivados de críticas literarias y poéticas como también de investigaciones psicolingüísticas y cognitivo experimentales. Las medidas se basan en las características del lenguaje expresivo y evocativo y su significación en diferentes situaciones del discurso. La *Concreción* es un reflejo de la calidad perceptiva y sensorial del sujeto, y se refiere a la medida en que las expresiones verbales hacen referencia a las propiedades sensoriales de las cosas o sucesos reales, o a cualquier cosa que sea experimentada a través de las sensaciones o el sentimiento. Incluye también la referencia a imágenes en cualquier modalidad sensorial (Bucci, 1993). En cuanto a la *Especificidad*, un texto altamente específico es detallado e informativo, hace referencia a objetos, personas, lugares y épocas en particular; especifica cantidades precisas y describe detalladamente al sujeto o al objeto del discurso. La dimensión de la *Claridad* debe ser entendida como reflejo de la conciencia que tiene el orador del intento comunicativo del discurso; hace referencia al grado de conciencia que tiene el orador de la perspectiva del oyente y del intento de llevarle a ese ámbito. La *presentación en imágenes* hace referencia a la impresión general que produce el lenguaje del orador en cuanto a la viveza y efectividad con que es capaz de reflejar y captar las imágenes o experiencias emocionales en cualquier modalidad sensorial.

En los trabajos iniciales de medición de AR, se utilizaron las fluctuaciones de AR para definir el ciclo referencial. Esto se realizó teniendo en cuenta que un discurso con AR alta es concreto, claro y específico considerándolo entonces una medida directa de las conexiones referenciales. Por otro lado, un discurso de AR baja es general, vago y utiliza lenguaje más abstracto reflejando así conexiones referenciales más débiles entre los sistemas de la mente (Rosow, 1997).

La AR puede ser calificada de una manera confiable por jueces, empleando escalas derivadas conceptualmente de los indicadores de los procesos referenciales definidos por Bucci (1997, 2002). En este sistema de evaluación es necesario dividir el texto en unidades de ideas y escalas de valoración. En la evaluación de la AR con jueces humanos se emplean dos tipos de segmentación: Unidades Temáticas Principales (UTP) y Unidades de Ideas (UI).

Las medidas de AR han sido ampliamente utilizadas en investigaciones clínicas y psicolingüísticas. La dimensión de AR ha mostrado ser operativa en funciones como en la autorregulación de emociones, en la comunicación interpersonal, en la capacidad de síntesis de la experiencia cognitiva, lingüística y emocional. Detalles concernientes a la aplicación de las medidas y de su confiabilidad y validez pueden ser halladas en el trabajo de Bucci (1997, 2002).

Además de las valoraciones mediante jueces, se han utilizado, en la investigación del ciclo referencial, análisis de contenido asistidos por computadora (Rosow, 1997).

La Actividad Referencial medida por computadora (CRA) es una medida de lenguaje computarizado que emite una representación gráfica de los ciclos referenciales del sujeto tal y como se manifiestan en las sesiones terapéuticas (Mergenthaler, 1996; Bucci, 1993). Permite la detección de esquemas emocionales en el transcurso del proceso terapéutico (Sammons y Seigel, 1998). El primer método para calificar el AR en forma computacional en inglés es el «Computerized Referential Activity» (CRA) de Mergenthaler y Bucci (1999).

El diccionario de CRA fue construido empíricamente a partir de varios pasos, basándose en las escalas de AR. Primero, jueces expertos calificaron diversos textos que incluyeron protocolos del TAT y memorias tempranas. Estos ítems se ordenaron de menor a mayor según su calificación de Actividad Referencial. Se retiraron aquellos del primer tercio y del último tercio, Los del tercio del medio fueron descartados. Así se creó una lista de alta Actividad Referencial y una lista de baja Actividad Referencial. Luego, se removieron todas las palabras cuya frecuencia en los textos era menor al 1% y aquellas cuyo dominio era muy específico. A partir de esto se creó un vocabulario característico que se puede aplicar a una gran variedad de textos.

Una vez establecido el primer vocabulario característico el siguiente paso fue aplicarlo utilizando el programa TAS a la muestra original para descubrir y eliminar aquellas palabras que no distinguían el grupo de alta AR del de baja. Este paso se lo repitió para diferentes muestras de textos hasta encontrar una lista final representativa (Rosow, 1997).

Por lo tanto, el CRA se computa en base a dos listas de palabras separadas: a) Alta Actividad Referencial, que consiste en el tipo de palabras que la gente probablemente use para describir imágenes y eventos concretos y específicos; b) Baja Actividad Referencial, integrada por palabras asociadas con mayor abstracción y lenguaje general. Finalmente, se computan ambas escalas por separado y luego se deriva el nivel de Actividad Referencial como diferencia entre los dos puntajes (Roussos, Acosta, Juárez y Mergenthaler, 2001).

En general, el CRA ha sido aplicado utilizando uno o dos sistemas de análisis de texto, el TAS/C basado en UNIX (Mergenthaler, 1985) y el CM basado en Windows (Mergenthaler, 1998). Ninguno permite el uso de diccionarios ponderados y ambos requieren una segmentación del texto en unidades arbitrarias por bloques de palabras, ya sea por conteo de palabras de longitud finita como en el caso de jueces humanos.

La necesidad de generar un nuevo diccionario para la evaluación de la Actividad Referencial surge a partir de la aparición de un nuevo programa para el análisis de texto asistido por computadora, el «Discourse Attribute Analysis Program» (DAAP) (Maskit, Bucci & Roussos, 2001). Este nuevo programa de análisis cuantitativo del discurso, posibilita la representación numérica y gráfica de distintos estilos lingüísticos, utilizando como unidad de análisis la palabra. Hasta la aparición de este nuevo programa, era necesario segmentar de distintas formas los textos, ya sea estableciendo unidades de sentido previas o segmentando por cantidad de palabras fijas. Maskit, Bucci y Roussos (2001), generaron un programa que analiza los textos sin segmentarlos. Esto se realiza a partir del establecimiento de un promedio ponderado móvil para cada palabra y de esta manera es posible representar en forma fidedigna cada punto del texto. A rasgos generales, el DAAP primero lee cada palabra en cada giro del discurso o unidad del mismo, una palabra a la vez, y compara esa palabra con las del diccionario. Después, computa usando la concordancia con palabras del diccionario o los valores para las palabras vecinas en esa unidad del discurso y con esto realiza un promedio ponderado móvil de los pesos del diccionario. La función resultante (con el número de palabra como variable independiente) es una aproximación numérica a una función matemática suavizada que aproxima el cambio de densidad de la variable psicológica o lingüística medida (Maskit, Bucci y Roussos, 2001).

Esta nueva generación de programas necesita a su vez, de nuevos diccionarios diseñados para trabajar sin la segmentación de los textos. Esto se debe a que los diccionarios por segmentación requieren un número mínimo de palabras por segmento para realizar el análisis (se basan principalmente en la redundancia de términos), mientras que el DAAP toma todo el texto y la densidad de palabras es sólo considerada a partir del texto en su totalidad. Por este motivo, se comenzó a diseñar un diccionario para la medición de la Actividad Referencial que se adecuara a la técnica de ponderación de los promedios utilizada por el DAAP (Waizman y Roussos, 2004). Este nuevo diccionario fue bautizado por Bucci y Maskit como «Diccionario de Actividad Referencial Ponderado» (Bucci y Maskit, 2004).

La construcción de versión en español de este diccionario de la Actividad Referencial ponderado (WRAD) es uno de los objetivos de este trabajo.

Métodos

Muestra

Para la construcción de un diccionario ponderado de la Actividad Referencial en español (S-WRAD) se analizaron material clínico y cuasi-clínico transcritos.

Se seleccionó la mayor variedad de material clínico posible con el fin de obtener un diccionario que pueda ser utilizado no solamente para una región de habla hispana sino para todas ellas, todos los sexos, edades, niveles de educación, etc. Es por esto que el material incluye (ver tabla 1 y 2):

- Ambos sexos.
- Diferentes estados civiles.
- Personas de distintos países de habla hispana: 16 textos argentinos, cinco cubanos, uno español y uno uruguayo.
- Diferentes niveles educativos (primario, secundario y universitario).
- Diferentes marcos teóricos de los terapeutas en las sesiones evaluadas.
- Diferentes tipos de materiales.

Estos materiales clínicos y cuasi clínicos fueron grabados de distintas díadas paciente-terapeuta y de entrevistas.

Las edades de los entrevistados varían entre 19 y 64 años.

Los textos S1 y S11 se utilizaron para evaluar el acuerdo interjueces de la calificación de la AR.

A partir de los textos S2, S3, S4, S5, S12, S16, S17, S19, S20, S21, S22, S23, S24 y S25 se construyó el diccionario.

Los textos S13, S14 y S15 se emplearon para la elección del parámetro de peso P.

Utilizando los textos S1, S6, S7, S8, S9 se probó el funcionamiento del diccionario una vez construido y se comparó con los puntajes obtenidos de AR calificados por jueces.

La información básica del conjunto de sesiones utilizadas se resume en las tablas 1 y 2.

Tabla 1.- Esquema de las características de las sesiones utilizadas

CODIGO	NOMBRE DEL CASO	NACIONALIDAD	EDAD	GÉNERO	TIPO DE MATERIAL	NIVEL DE EDUCACIÓN	ESTADO CIVIL	MARCO TEÓRICO
S1	Paula	Argentina	19	F	Sesión terapéutica	Secundario	Soltera	Sistémico
S2	Luz	Argentina	21	M	Sesión terapéutica	Universitario	Soltera	Cognitivo
S3	Maru	Argentina	19	F	Sesión terapéutica	Universitario	Soltera	Psicoanalítico
S4	José	Cubana	30-40	M	Memoria temprana	Universitario	Desconocido	---
S5	Manuela	Española	24	F	Sesión terapéutica	Universitario	Desconocido	Psicoanalítico
S6	Mrs. C	Traducido		F	Sesión terapéutica	Universitario	Casada	Psicoanalítico
S7	Osvaldo	Uruguay	41	M	Entrevista	Desconocido	Desconocido	---
S8	Toto	Argentina	35	M	RAP	Universitario	Desconocido	---
S9	Ana	Argentina	50	F	RAP	Primario	Casada	---
S10	Tito	Argentina	19-30	M	RAP	Secundario	Desconocido	---
S11	Lidia	Argentina	52	F	Memoria temprana	Universitario	Casada	---
S12	Martin	Cubana	30-40	M	Memoria temprana	Universitario	Desconocido	---
S13	Ariel		30-40	M	Memoria temprana	Desconocido	Desconocido	---
S14	Samantha	Cubana	30-40	F	Entrevista	Desconocido	Desconocido	---
S15	Marta	Cubana	30-40	F	Entrevista	Desconocido	Desconocido	---
S16	Alfredo	Argentina	30-40	M	Sesión terapéutica	Desconocido	Desconocido	Psicoanalítico
S17	Andrea	Argentina	30-40	F	Sesión terapéutica	Desconocido	Casada	Psicoanalítico
S18	Mariana	Argentina	42	F	Entrevista	Universitario	Casada	---
S19	Juanita	Argentina	23	F	Entrevista	Universitario	Soltera	---
S20	Elba	Argentina	60	F	Entrevista	Universitario	Casada	---
S21	Elbo	Argentina	64	M	Entrevista	Secundario	Casado	---
S22	Claudia	Argentina	22	F	Entrevista	Universitario	Soltera	---
S23	Edgardo	Argentina	24	M	Sesión terapéutica	Secundario	Soltero	Psicoanalítico
S24	Aurora	Argentina	54	F	Entrevista	Universitario	Casada	---
S25	Francisco	Cubana	30-40	M	Memoria temprana	Desconocido	Desconocido	---

Tabla 2.- Síntesis del material utilizado

NOMBRE	CATEGORÍA	NÚMERO
Tipo de material	Sesión terapéutica	7
	Entrevista	9
	RAP	3
	Memoria temprana	5
Género	Femenino	14
	Masculino	10
Estudios	Primario	1
	Secundario	4
	Universitario	11
	Desconocido	8
Edad	19 - 30	8
	30 – 40	9
	Más de 40	7
Estado civil	Soltero	6
	Casado	7
	Desconocido	11
Nacionalidad	Argentina	16
	Uruguay	1
	Española	1
	Cubana	6

Materiales

Los tipos de materiales utilizados fueron:

- sesiones clínicas,
- entrevistas a terapeutas,
- entrevistas RAP (Paradigma anecdótico relacional, Luborsky, 1997),
- entrevistas de memorias tempranas.

Para la construcción del diccionario la autora diseñó un programa especialmente formulado para recrear los pasos definidos en el trabajo original de Bucci y Maskit (2004). Este programa consta de una serie de rutinas con el fin de contabilizar la frecuencia de los distintos ítems léxicos y el armado de los pesos de los mismos. Dicho programa fue generado en el paquete estadístico S-plus (2000).

Para la aplicación del diccionario a nuevos textos, su posible análisis y comparación con la AR, se utilizó otro programa creado por la autora especialmente para tal fin y el DAAP (Maskit, Bucci y Roussos, en prensa), mencionado en la introducción. El programa creado para el análisis de nuevos textos lo reconoce las distintas palabras y les asigna los pesos del diccionario, si no aparece en el diccionario le concede ponderación cero. Luego realiza, para cada unidad temática, la suma y el promedio de las ponderaciones para poder compararlos con las calificaciones de los jueces.

Todos los códigos de los programas especialmente confeccionados para esta tesina se adjuntan en el apéndice 1.

Procedimiento

Paso 1: Calificación de RA

Para la construcción del diccionario, se entrenaron tres jueces en la evaluación de la Actividad Referencial. Este entrenamiento y su posterior evaluación, fue llevada a cabo por la Lic. Waizman (en preparación), integrante del grupo de investigación. Se utilizan los resultados obtenidos.

Primero, los jueces fueron entrenados para la segmentación de los textos. Fue necesario dividir los textos en unidades temáticas o unidades de ideas. Las unidades de ideas se asocian con cambios en los

temas del discurso. Se usan para facilitar la calificación de textos largos como monólogos empleados en estudios experimentales.

Las unidades de ideas reflejan cambios en el foco de la narrativa del discurso, que pueden ser menos obvias que los temas y su tamaño más adecuado para la calificación de AR. Se las puede definir como cambios en el humor o sentimiento de un pasaje, la introducción de imaginación, como cambios en la escena o persona.

La segunda etapa fue la calificación mediante las escalas de AR.

Los jueces fueron entrenados de modo tal de no efectuar inferencias clínicas sobre el material y mantener el foco en los contenidos manifiestos del texto.

Luego de analizar la definición de cada escala de AR, los tres jueces evaluaron la misma sesión terapéutica (S1) con el fin de calcular el acuerdo interjueces.

Un juez norteamericano bilingüe evaluó los mismos textos y se calculó el acuerdo interjueces comparándolo con los otros jueces. Después de analizar el acuerdo y reentrenar a los jueces de modo tal de obtener un mayor grado de acuerdo, los tres evaluaron una entrevista (S11).

Luego de estudiar los resultados se creó una lista de ejemplos de calificaciones con el fin de tener una guía para la evaluación.

Todos los textos fueron calificados por jueces siguiendo el manual de calificación de AR de Bucci.

Cada uno de los jueces calificó el texto en una escala de 0 a 10. El puntaje final para cada juez fue el promedio entre de las cuatro escalas de Actividad Referencial, y el puntaje final utilizado para el procesamiento fue el resultante de la evaluación del acuerdo interjueces. Se dividieron las distintas sesiones y cada una fue calificada.

Paso 2: División de la AR

El puntaje de AR fue asignado por jueces en una escala de números naturales del 0 al 10. El puntaje promedio final del AR fue redondeado para cada texto a la 1/8 (0,125) más cercano, dejando 81 divisiones en la escala de 0 a 10. Los 81 puntajes se dividieron para formar 6 categorías.

La tabla 3 describe cada una de las 6 categorías.

Estas divisiones fueron elegidas respetando las realizadas en la construcción de WRAD (Bucci y Mas-Kit, 2004) para poder compararlos.

Tabla 3.- Categorías de la AR utilizadas

CATEGORÍA DE AR	RANGO DE PUNTAJE DE AR	NUMERO DE ITEMS LÉXICOS
Muy baja	0 – 2.75	5898
Moderadamente baja	2.875 – 3.75	8582
Levemente baja	3.875 – 4.75	6527
Levemente alta	4.875 – 5.75	5626
Moderadamente alta	5.875 – 7.125	12952
Muy alta	7.25 – 10	10688
Total	0 – 10	50323

Paso 3: La matriz principal

La autora creó un programa de computación en el paquete estadístico S-Plus (ver apéndice 1) con el fin de detectar los distintos ítems léxicos presentes en las sesiones utilizadas para la construcción del diccionario.

Lo primero que realiza dicho programa es distinguir los diferentes ítems léxicos de los textos y contabilizar su frecuencia total.

Una vez discriminados los diferentes ítems léxicos presentes en estos textos, se procedió a contabilizar la frecuencia de cada uno en los distintos segmentos de texto calificados por los jueces. Este procedimiento también se realizó mediante un programa realizado para esta tarea.

Utilizando esta información, se construyó una primera matriz. Para cada fila, la primera columna contiene el ítem léxico y el resto de las 6 columnas restantes contienen el número de veces que se encontró ese ítem léxico en cada una de las categorías de RA. La matriz principal de 470 filas y 8 columnas (apéndice 2) se obtuvo de esta primera matriz usando el siguiente procedimiento:

- Primero, cualquier ítem cuyo número total de aparición en el texto fuese menor a 11 se descartó. Este procedimiento de eliminación hizo que las filas se redujeran a 470.
- Dado que el número total de ítems léxicos en cada categoría no es uniforme, las entradas en cada categoría se las cambió a proporciones, es decir el número de veces que un ítem estaba en una categoría dividido el número de total de ítems de esa categoría (última columna de la tabla 3). La última columna muestra el número total de veces que aparece cada ítem léxico en el total de las sesiones.

A modo ilustrativo, se muestran en la tabla 4 las primeras 20 filas ordenadas según la frecuencia de aparición de la tabla principal. La última columna de la tabla 4 indica el peso que tiene cada ítem en el diccionario S-WRAD. La manera de obtener este peso se explica más adelante.

Tabla 4.- las 20 primeras filas de la matriz principal

Palabra diccionario	Categoría 1	Categoría 2	Categoría 3	Categoría 4	Categoría 5	Categoría 6	Frecuencia total	Ponderación para P=0.95
que	0.0705	0.0665	0.0729	0.0704	0.0640	0.0641	3374	NA
no	0.0493	0.0627	0.0340	0.0297	0.0401	0.0393	2157	-0.625
de	0.0322	0.0345	0.0467	0.0395	0.0437	0.0398	2005	-0.25
y	0.0331	0.0368	0.0345	0.0437	0.0400	0.0432	2002	NA
a	0.0251	0.0271	0.0323	0.0313	0.0302	0.0297	1477	-0.125
la	0.0229	0.0225	0.0326	0.0325	0.0276	0.0325	1430	NA
en	0.0210	0.0209	0.0280	0.0258	0.0223	0.0250	1187	-0.25
me	0.0219	0.0241	0.0207	0.0213	0.0236	0.0253	1168	NA
el	0.0185	0.0200	0.0248	0.0163	0.0217	0.0245	1078	NA
es	0.0295	0.0248	0.0195	0.0146	0.0170	0.0178	1006	-1
lo	0.0239	0.0153	0.0165	0.0165	0.0198	0.0177	918	-1
yo	0.0212	0.0177	0.0167	0.0202	0.0173	0.0182	911	-1
un	0.0114	0.0108	0.0165	0.0160	0.0145	0.0157	714	-0.125
se	0.0122	0.0108	0.0089	0.0139	0.0153	0.0172	683	1
con	0.0093	0.0147	0.0159	0.0146	0.0119	0.0139	670	-0.25
como	0.0168	0.0176	0.0121	0.0116	0.0109	0.0121	664	-0.75
por	0.0115	0.0112	0.0116	0.0130	0.0107	0.0163	626	1
una	0.0105	0.0115	0.0106	0.0126	0.0132	0.0129	610	0.625
mi	0.0059	0.0094	0.0104	0.0167	0.0137	0.0117	600	0.25
si	0.0117	0.0113	0.0144	0.0087	0.0124	0.0110	587	-0.25

Paso 4: Método para la asignación de la ponderación dependiendo de un parámetro

Para cada uno de los ítems léxicos una de las 6 proporciones de las categorías de AR es necesariamente mayor que las otras. A esta categoría se la llama *categoría dominante*. Para establecer cuán fina uno quiere que sea la discriminación de las categorías, se introduce un **parámetro de peso P**, cuya elección se explica más adelante. Este parámetro P provee una medida de cuanto más grande es esta proporción dominante del resto de las proporciones. La asignación de un peso a cada ítem léxico depende del nivel de P que se elige según el criterio de dominancia. El siguiente procedimiento describe cómo se asigna cada peso a los distintos ítems léxicos una vez que se obtuvo P.

Primero se identifica la categoría dominante para cada ítem léxico que es aquella que posee la proporción máxima. A esta proporción máxima se la llama M, y a la categoría en la que ocurre C. Luego se calcula $P \cdot M$ usando el valor seleccionado de P. Si los valores en las otras 6 columnas son menores que esta multiplicación, entonces ese ítem va a ser incorporado al diccionario con el peso correspondiente a la categoría dominante según el esquema de la tabla 5. Los pesos varían entre -1 y +1, representando las 6 categorías de muy baja a muy alta RA. La forma de asignación de la ponderación a cada categoría es la misma que la utilizada en la construcción del WRAD por Bucci y Maskit (2004).

Tabla 5.- Ponderación para cada categoría dominante absoluta

CATEGORÍA DE AR	PONDERACIÓN
1	-1,0
2	-0,625
3	-0,25
4	0,25
5	0,625
6	1

Si este primer paso falla en proporcionar una entrada, o sea que al hacer $P*M$ el resto de las categorías no son menores, entonces se observan las celdas cuyo producto no dio menor. Si una de estas celdas no es adyacente a la dominante, entonces se considera que este ítem tiene distribución bimodal y se descarta del diccionario. Si se encuentran dos celdas adyacentes a la dominante una en cada lado entonces el peso que se asigna es el mismo que el mencionado antes. Si se encuentra una celda adyacente a la dominante, el peso se asigna según la tabla 6.

Tabla 6.- Asignación de la ponderación en el caso de dos categorías dominantes.

CATEGORÍA DOMINANTE	CATEGORÍA CON EL SIGUIENTE VALOR MÁS ALTO	PONDERACIÓN
1	2	-0,875
2	1	-0,75
2	3	-0,5
3	2	-0,375
3	4	-0,125
4	3	0,125
4	5	0,375
5	4	0,5
5	6	0,75
6	5	0,875

Paso 5: Selección del parámetro de peso P

Los textos S13, S14 y S15, que no se utilizaron para la selección de los ítems del diccionario y su cálculo de la ponderación, se emplearon para encontrar el mejor valor del parámetro de peso P.

También, se usaron los textos de la construcción del diccionario para calcular otro parámetro de peso P y compararlo con el anterior.

Se puede construir un diccionario ponderado WRAD (P) para cada parámetro P utilizando el procedimiento anterior.

Para cada diccionario WRAD(P) se puede obtener un puntaje para cada texto de los distintos conjuntos, como la suma de los pesos para cada ítem en el texto que se corresponde con un ítem en el diccionario. Para obtener este puntaje, cada ítem en el texto se lo compara con los ítems del diccionario. Si el ítem se corresponde con uno del diccionario, entonces el puntaje del WRAD sube (baja) el valor positivo (negativo)

del peso asignado a ese ítem en el diccionario. Si el ítem del texto no se corresponde con ningún ítem en el diccionario, entonces el puntaje no aumenta ni disminuye. Una vez que se compararon todas las palabras del texto, el puntaje total del texto es la suma de todos los pesos. Debido a que hay una fuerte correlación entre la longitud del texto y el puntaje total de RA, se deben realizar distintos procedimientos para abordar este problema. El método seleccionado en este trabajo fue corregir el puntaje de S-WRAD (P) según la longitud del texto de la siguiente manera: se introduce el puntaje promedio S-MWRAD (P) que se obtiene al dividir el puntaje S-WRAD por el número total de palabras del texto. En este trabajo se muestran los dos puntajes el S-WRAD y el S-MWRAD.

Usando un programa armado para realizar el procedimiento antes explicado, se calculó el puntaje S-WRAD (P) y el S-MWRAD (P) y se comparó la validez de cada S-WRAD (P) para los distintos P de la siguiente manera:

- Primero, se graficaron los valores obtenidos por el diccionario versus los valores dados por los jueces para las distintas unidades del texto.
- En función del gráfico, se calculó la correlación debido a que esta es la metodología utilizada en el WRAD (una mayor discusión sobre la comparación de resultados a través del uso de correlaciones se presenta en la sección de discusión y conclusiones).
- Este procedimiento fue realizado con el conjunto de textos de la construcción del diccionario S-WRAD (P) y con el conjunto de textos de selección del P utilizando valores de P entre 0,80 y 0,99 en pasos de 0,01.

Se observa que la correlación aumenta desde $P = 0,80$ hasta $P = 0,95$ y luego disminuye. El valor de la correlación más alta ocurre en $P = 0,95$ y es de 0,59 para el S-MWRAD. El diccionario final se lo calculó utilizando este P.

Paso 6: Ajustes finales

Todos los ítems del diccionario fueron revisados y se encontraron anomalías que fueron corregidas. Se eliminaron los nombres propios incluyendo nombres de días dado que en principio no son indicadores de un nivel de Actividad Referencial específico. A todos los números que aparecieron en el diccionario se les asignó un peso de +1.

Una vez hechas las modificaciones antes mencionadas siguiendo los pasos del WRAD, se obtuvo el diccionario final S-WRAD en castellano con 470 ítems.

El último paso en el procedimiento fue probar el diccionario S-WRAD en nuevos textos, esto quiere decir que se obtuvieron puntajes de la suma de los ítems en cada unidad de texto (S-WRAD) y del promedio (MS-WRAD) para cada texto que luego fueron comparados con puntajes de AR dados por los jueces.

Resultados

Descripción de WRAD en castellano

El diccionario S-WRAD es una lista de palabras que cuenta con 470 ítems léxicos. Dentro de los ítems de este diccionario, 467 palabras son ordinarias y tres corresponden a interjecciones (eh, mj y je).

Es interesante destacar que en el diccionario aparecen todas las palabras utilizadas para nombrar vínculos familiares como hermano, padres, tíos, abuelos e hijos. Por otro lado, también incluye a las palabras analista y terapeuta.

La palabra más frecuente resultó ser «que». Entre las 20 palabras más frecuentes se encuentra «no» con una ponderación negativa para la Actividad Referencial.

Dentro de las 20 palabras más frecuentes una sola es un verbo, no hay sustantivos ni adjetivos, cinco son artículos, siete son conjunciones, cinco preposiciones y el resto pronombres.

De las palabras más frecuentes que se muestran en la lista de la tabla 7 del SWRAD, tres son bimodales esto quiere decir que hay dos proporciones máximas no consecutivas; y el 70% de las restantes tiene ponderación negativa.

Comparación con el diccionario en inglés

Tabla 7.- Las 20 primeras palabras más frecuentes en castellano y en inglés.

SWRAD	PONDERACIÓN	WRAD	PONDERACIÓN
que	NA	I	0.75
no	-0.625	and	1
de	-0.25	The	1
y	NA	To	0.25
a	-0.125	it	0.875
la	NA	s	-1
en	-0.25	She	1
me	NA	That	-0.875
el	NA	A	0.625
es	-1	Was	1
lo	-1	t	-0.625
yo	-1	of	-0.625
un	-0.125	in	1
se	1	Mm	-0.625
con	-0.25	He	0.625
como	-0.75	You	-0.625
por	1	Her	1
una	0.625	My	0.625
mi	0.25	Is	-1
si	-0.25	They	0.625

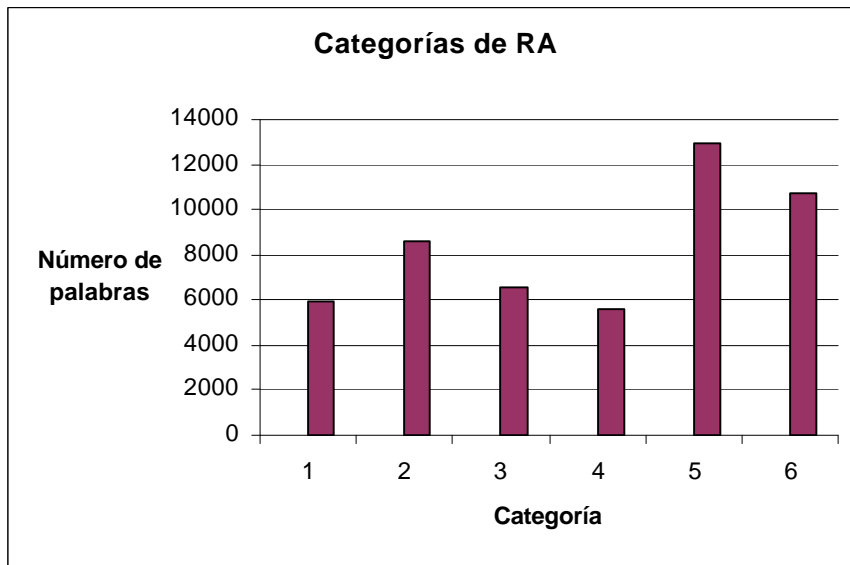
La palabra «yo» aparece tanto en el diccionario en inglés como en castellano dentro de las 20 palabras más frecuentes; pero con ponderaciones distintas. Además, hay otras palabras que se repiten en ambos diccionarios como: «mi», «a», «y», «la» o «el» por «the», «en», «es». Esto implica que casi un 50% de las palabras más frecuentes coinciden en ambos idiomas. De hecho, del resto de las palabras que no coinciden, varias corresponden a aquellas que no tienen traducción directa. Esto estaría mostrando una universalidad en las palabras de uso más frecuente y en la Actividad Referencial.

Análisis del diccionario

Categorías de RA

Como se observa en la tabla 3 las categorías tomadas no son parejas respecto a la cantidad de palabras que albergan. Esta información se muestra en forma gráfica en la figura 1.

Figura 1.- Cantidad de palabras en cada categoría de RA.

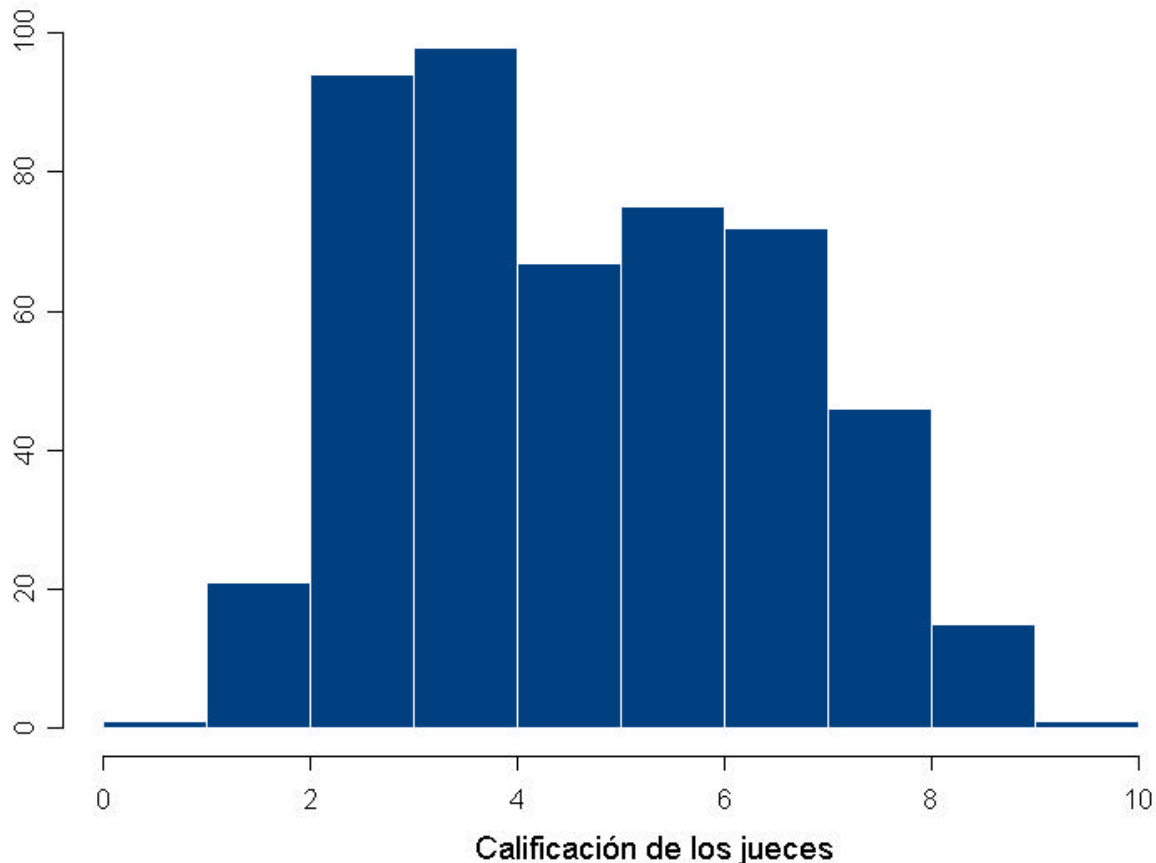


Las categorías tomadas fueron las mismas que se crearon para el WRAD, de modo tal de que sean parejas para esa muestra. Para explorar la causa del hecho de que las categorías sean desparejas en castellano, se analizaron dos variantes:

- Calificación de los jueces.
- Cantidad de palabras en las unidades calificadas por los jueces.

Respecto de la calificación de los jueces, se exploró si había alguna tendencia en los mismos a calificar con Actividad Referencial alta lo que explicaría el hecho de que estas categorías fuesen las de mayor número de palabras. Para esto se muestra en la figura 2 un histograma con de las distintas calificaciones de los jueces.

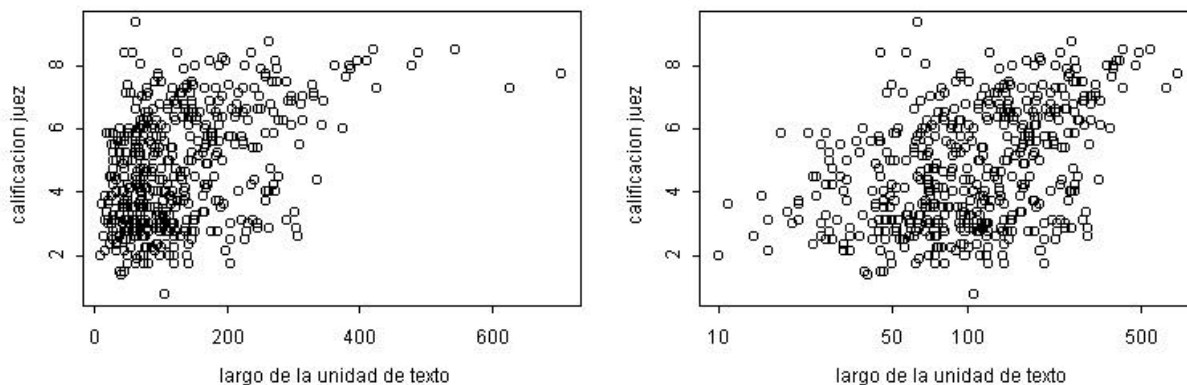
Figura 2.- Cantidad de veces que los jueces calificaron con los distintos puntajes



Se observa a partir de este gráfico, que no hay una tendencia de los jueces a calificar con puntajes altos. Es más, hay una leve tendencia a calificar con puntajes bajos. Por lo tanto, la variante de la calificación de los jueces no explicaría la mayor cantidad de palabras en las categorías más altas.

Otra causa posible es que las unidades de los textos calificadas con Actividad Referencial alta tuviesen mayor número de palabras. Para analizar esta hipótesis se graficó la cantidad de palabras versus la calificación de los jueces que se muestra en la figura 3.

Figura 3.- Cantidad de palabras de la unidad versus calificación de los jueces en escala decimal y en escala logarítmica.



En este caso sí se puede observar una relación entre la longitud de la unidad del texto y la calificación de los jueces. Las calificaciones más altas corresponden a unidades con más palabras. Este resultado es esperable debido a que, a medida que la Actividad Referencial es más alta, se utilizan más palabras, dado que se intenta ser más preciso, más concreto, etc.

Pero este resultado no es el único que explicaría que las categorías sean desparejas en cuanto al número de palabras. Al estudiar en detalle los rangos de las mismas se observan que no son iguales, las categorías más altas tienen rangos mayores.

Por lo tanto, la existencia de mayor cantidad de palabras en las unidades calificadas con Actividad Referencial alta y la diferencia en los rangos de las categorías explicarían el hecho de que las distintas categorías no alberguen un número parejo de palabras.

Correlación interna

Se calculó la correlación de Pearson entre los valores del diccionario y la calificación de los jueces sobre todos los archivos utilizados para la construcción del diccionario.

Este procedimiento de comparación se realizó para distintos P y los resultados obtenidos se muestran en la figura 4. Se indica este resultado tanto para la suma como para el promedio.

Figura 4.- Correlación interna para los distintos P con la suma.

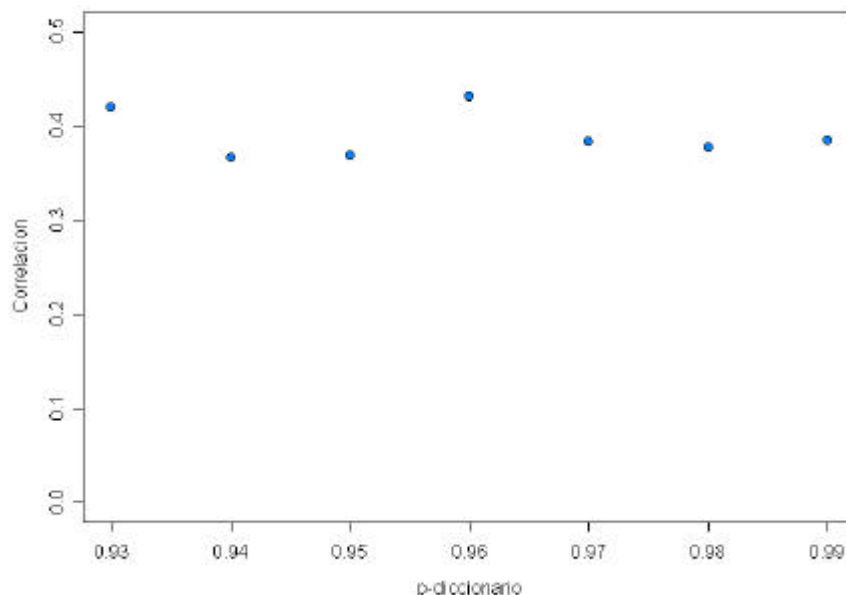
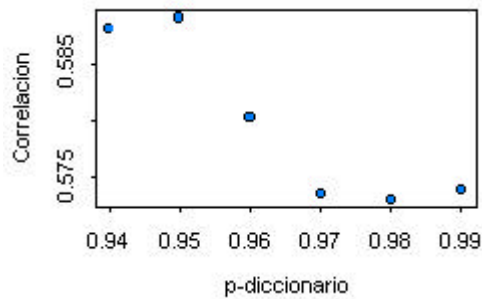


Figura 5.- Correlación interna para distintos P para el promedio MSWRAD.

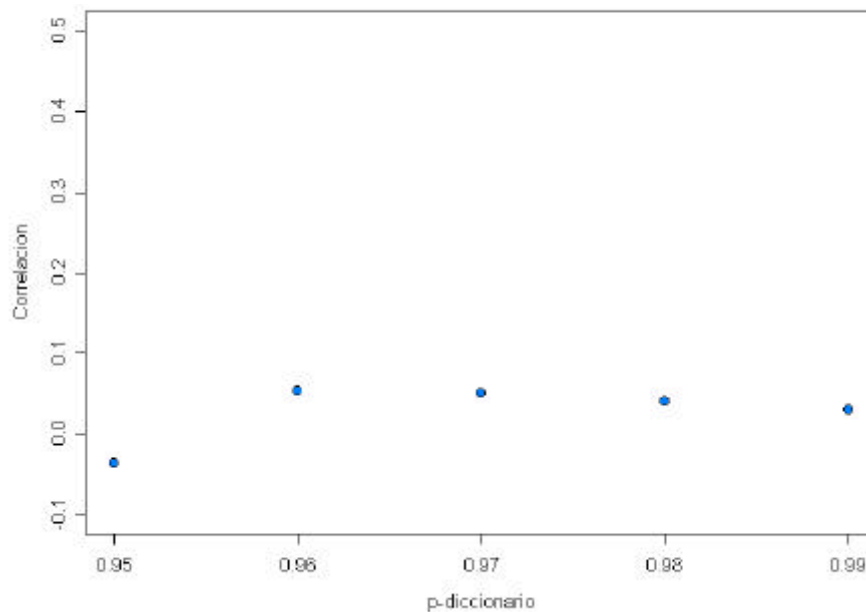
Dado que hay una dependencia con la longitud del texto, se toma el mejor P del análisis del promedio.

El mejor P resultante es $P = 0,95$ con una correlación de 0,59 para el promedio. Se elige el P de los resultados del promedio ya que para este caso se obtienen valores más altos.

Correlación de prueba

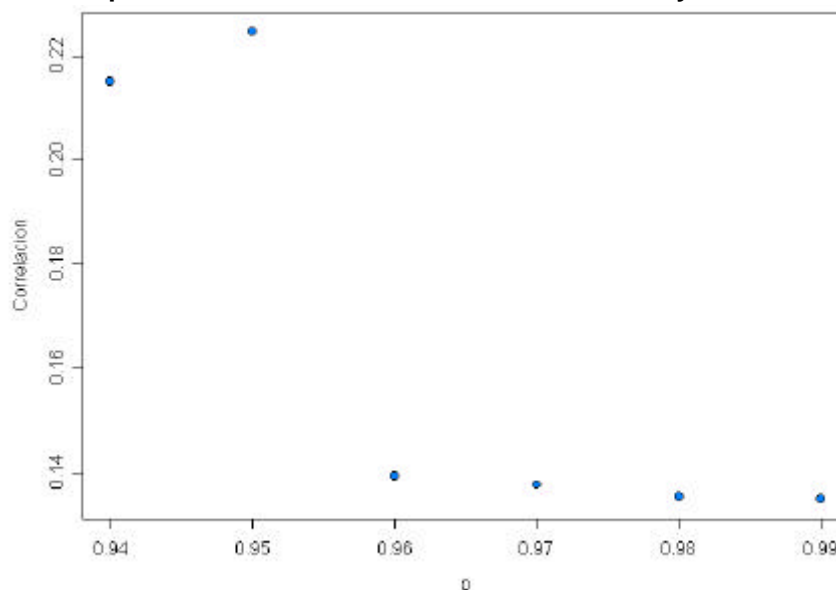
Se realizó la correlación entre la calificación de los jueces y los resultados del diccionario para los archivos S13, S14 y S15 que no se habían utilizado con el objeto de su construcción.

En la figura 6 se muestra la correlación obtenida para los distintos P.

Figura 6.- Correlación de prueba para la selección del parámetro de ponderación P.

Se observa que las correlaciones son muy bajas, no mayores a 0,1. Esto indicaría que, en este caso, no hay relación entre la calificación de los jueces y lo que se obtiene del diccionario. La explicación de este hecho se muestra en la siguiente sección.

Se descartaron estos archivos y se seleccionaron para calcular el P los textos S1 y S6. Se obtuvo a partir de ellos la figura 7.

Figura 7.- Correlación para la selección de P sobre los archivos S1 y S6.

Se observa en este caso que la correlación máxima de 0,22 ocurre para $P = 0,95$. Esto coincide con lo obtenido para la correlación interna, lo que es un resultado esperable. Por lo tanto para el diccionario final se eligió como parámetro de ponderación a $P = 0,95$.

Correlación externa

Se calculó la correlación de la calificación de los jueces y de los resultados obtenidos del diccionario para los archivos S7, S8 y S9 para el diccionario final con un parámetro de ponderación de 0,95. La correlación del promedio resultó de 0,505.

Cobertura

Se calculó para el conjunto de textos externo S7, S8 y S9 que no se utilizaron para construir el diccionario, el porcentaje de palabras que estaban incluidas en el diccionario. A esto se lo denominó cobertura. De este modo, la cobertura resultó del 70%.

Comparación con los resultados obtenidos para WRAD

Se muestra en la tabla 8 la comparación de los resultados obtenidos para el diccionario en inglés y para el diccionario en castellano.

Tabla 8.-Comparación de los resultados obtenidos en inglés y en castellano.

	WRAD	SWRAD
Correlación interna	0,55	0,585
Correlación de prueba	0,57	0,22
Correlación externa	0,38	0,505
Cobertura	85%	70%

Se observa que los resultados obtenidos para el diccionario construido en castellano son similares a los del inglés. Las correlaciones en ambos casos son del mismo orden.

Hay una diferencia en la cobertura entre el WRAD y el SWRAD. Un posible factor que explicaría esto es que el WRAD tiene 697 ítems léxicos mientras que el SWRAD tiene 470 ítems.

Lo significativo es que para el SWRAD se obtienen correlaciones similares al WRAD.

Discusión y conclusiones

Teniendo en cuenta los resultados presentados en la sección anterior, se puede concluir que el SWRAD construido, muestra resultados similares a los del WRAD publicados por Bucci y Maskit (2004). De este modo, se cumplió el objetivo de construir un diccionario de Actividad Referencial ponderado para idioma español utilizando los mismos procedimientos del diccionario presentado por Bucci y Maskit (2004).

Se realizó un análisis complementario sobre los resultados del diccionario y de la metodología empleada. Para ello, tal como se planteó en los objetivos específicos, se evaluó el método utilizado para analizar el acuerdo entre las dos formas de calificación de la AR.

Todos los resultados obtenidos se basaron en el uso de la técnica estadística de correlación. Dicha técnica fue utilizada en el caso del SWRAD debido a que se trataba de una réplica del diseño original. Lo que se buscaba al recurrir a la correlación era analizar el grado de acuerdo entre la calificación de los jueces y los resultados del diccionario.

La correlación tal como la define Altman (1991), «es un método de análisis que se emplea cuando se estudia la posible asociación entre dos variables continuas». Si se quiere medir el grado de asociación entre las dos variables se usa el coeficiente de correlación. El método estándar que es el empleado en ambos trabajos es el de la correlación de Pearson que lleva a una cantidad r que puede tomar valores entre -1 y 1 . **Este coeficiente r mide el grado de asociación lineal entre los valores de las dos variables.** Por lo tanto, se obtienen valores de $+1$ o de -1 si todos los puntos se encuentran en una línea recta perfecta. Una correlación de cero indica que no existe relación lineal entre los valores de las dos variables. En resumen, lo que mide r es una medida de la dispersión de los puntos alrededor de una línea recta (Altman, 1991).

Sin embargo, el interés en el caso del diseño de un diccionario de esta naturaleza lo que se busca es comparar el resultado de dos métodos distintos para medir una misma variable psicológica (la Actividad Referencial). Por lo tanto, lo que se buscaba era establecer el **grado de acuerdo** entre los métodos con el fin de mostrar que ambos eran intercambiables y que uno podía reemplazar al otro.

Tal como explica Altman (1991), los estudios como este, que realizan comparación de distintos métodos, son frecuentemente analizados erróneamente. Específicamente en muchos casos, se calcula la correlación entre los valores obtenidos por los dos métodos. Así se interpreta un alto valor de la correlación como un indicador de un buen acuerdo. Asimismo, hay diversas razones por las cuales el uso de la correlación es un análisis inapropiado. Primero, la correlación es una medida de la asociación lineal, como se explicó anteriormente, entre las variables, que no es lo mismo que una medida de acuerdo. El acuerdo se debe establecer en términos directamente relacionados con las mediciones, tanto en su magnitud como en las diferencias existentes entre medidas. No es posible interpretar un $r = 0,92$ en el mismo sentido de un límite de acuerdo. Segundo, se puede obtener un alto grado de correlación aun cuando el acuerdo es pobre (Altman, 1991).

En otras palabras, si dos métodos están de acuerdo linealmente, se obtiene una correlación alta. Pero no vale la inversa; el hecho de obtener una buena correlación no implica que dos métodos estén de acuerdo. Como veremos más adelante este podría ser el caso del diccionario construido. Un coeficiente de correlación bajo mostraría un bajo acuerdo entre los dos métodos. Esto quiere decir que hay una relación entre la correlación y el acuerdo lineal entre dos métodos pero que no se puede establecer el grado de acuerdo a través de la correlación.

Hay que destacar dos puntos en toda esta discusión sobre el acuerdo.

Si se toma una definición estricta sobre el concepto acuerdo, que implique que ambos métodos sean intercambiables debido a que miden lo mismo de distintas formas, se debe obtener como resultado valores similares. De esta forma, la gráfica ideal entre los valores de ambos métodos mostraría una recta a 45° . Por esto, es que correlación y acuerdo tienen una clara relación y a menudo, aunque erróneamente, se los utiliza en forma indistinta. En el caso estudiado en esta tesina, el acuerdo estricto era imposible de establecer dado que las escalas de medición de los jueces (0-10) y del diccionario (-1 y 1) son distintas y por lo tanto no se puede obtener valores parecidos.

Se puede ampliar la definición de acuerdo en el sentido de que un método pueda reemplazar al otro, dado que a partir de uno se pueda predecir con un buen grado de precisión lo que se obtendría por el otro método. En esta definición de acuerdo entrarían los cambios de escala y funciones inversibles de predicción. Bajo esta definición cabría analizar el acuerdo entre la calificación de los jueces y los resultados del diccionario. En cuanto a la correlación, en el caso de los cambios de escala, todavía podría seguir teniendo la relación antes descrita pero en el caso de otras funciones inversibles podría no existir relación alguna.

En conclusión, metodológicamente resulta inadecuado utilizar la correlación como medida de acuerdo entre dos métodos. Para analizar apropiadamente el acuerdo se deben emplear medidas específicas.

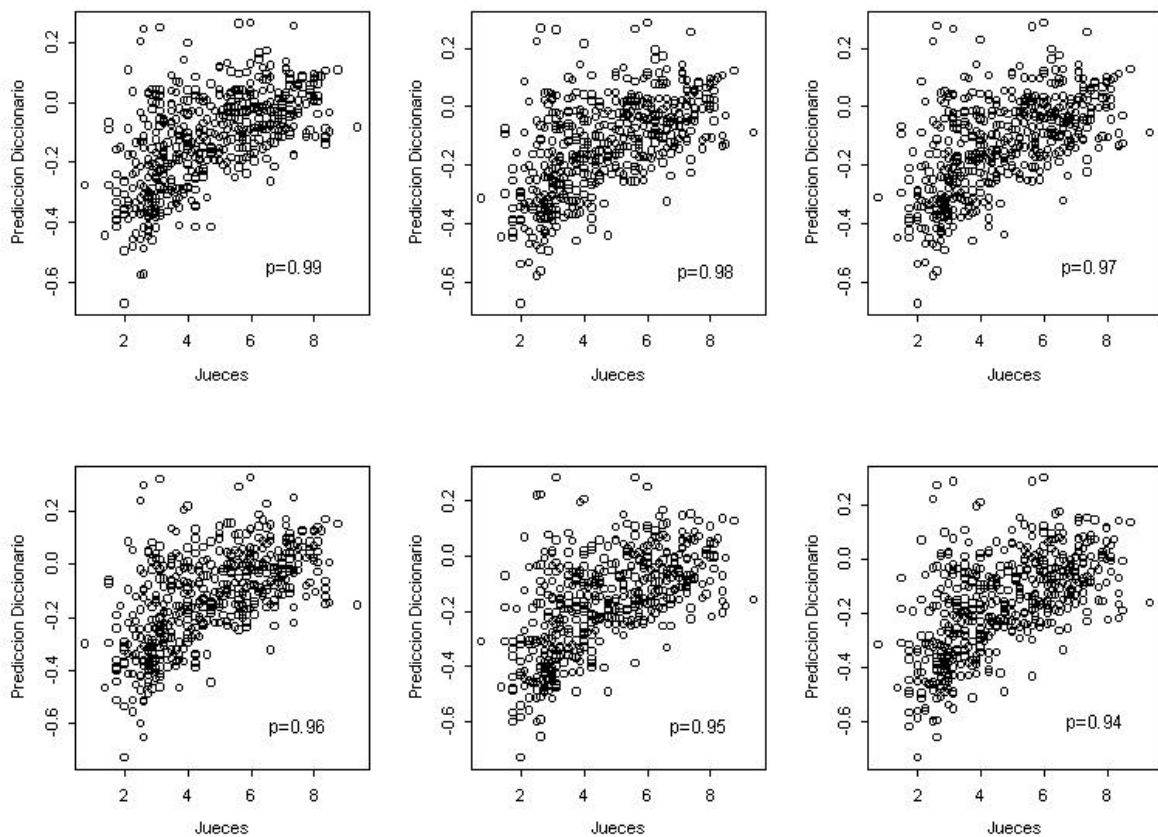
El otro punto que vale la pena destacar es que es posible medir el grado de acuerdo entre dos métodos cuando ambos métodos están midiendo la misma variable. Es decir, miden la misma operacionalización de un constructo. En el caso de esta tesina es válido preguntarse si la calificación de los jueces y el método del diccionario miden la misma variable o son distintas operacionalizaciones del mismo constructo psicológico. En este sentido, cuando se intenta modelizar estilos lingüísticos a través de diccionarios se debe definir y analizar apropiadamente la variable y su operacionalización con el fin de poder comparar resultados de distintos métodos.

Suponiendo que en este caso ambos métodos miden la misma variable, para establecer el acuerdo no se debe utilizar correlación sino otros métodos estadísticos de estimación específicos teniendo en cuenta que definición de acuerdo se está tomando.

Pero antes de establecer el acuerdo por cualquier método específico, es recomendable observar si hay una estructura en un gráfico de dispersión (scatter plot), o sea en un gráfico de puntos de un método versus el otro. Si no hay estructura en este gráfico, no tiene sentido calcular un acuerdo porque el resultado de esto puede ser engañoso.

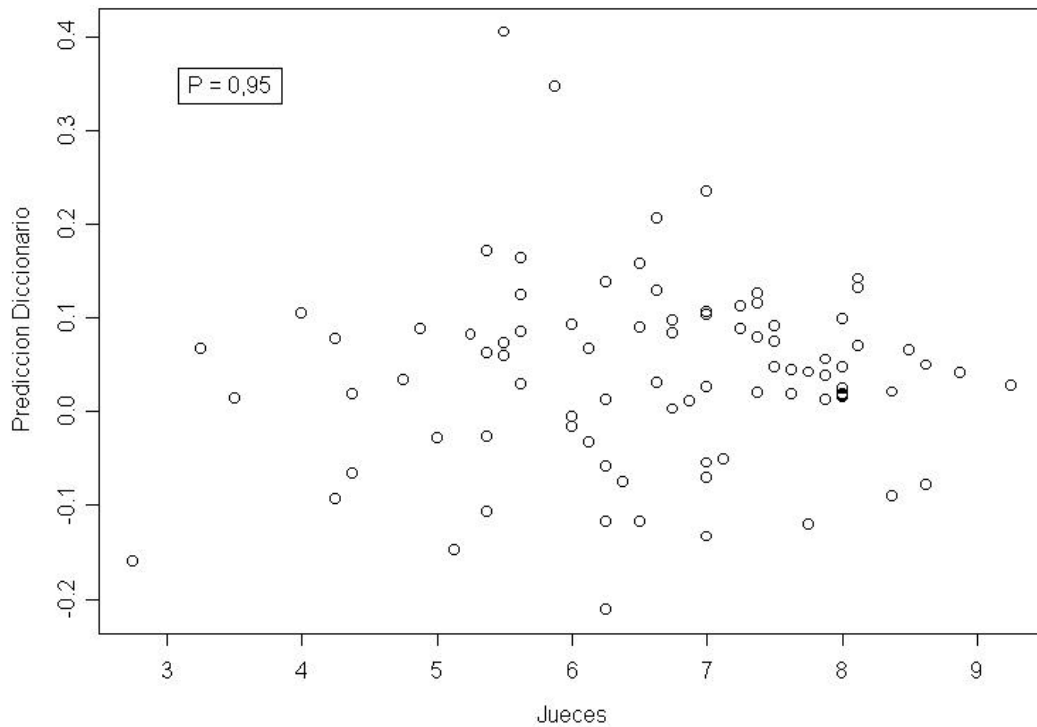
En el caso de esta tesina se muestran en las figuras 8, 9 y 10 los gráficos de dispersión entre ambos métodos en el caso interno, de prueba y externo, o sea con los tres conjuntos de textos. Se eligieron para mostrar los gráficos basados en el promedio, porque son más adecuados representar el cruce de ambos métodos.

Figura 8.- Gráficos de estructura para el promedio con el conjunto de textos utilizados para la construcción del diccionario para distintos parámetros de ponderación.



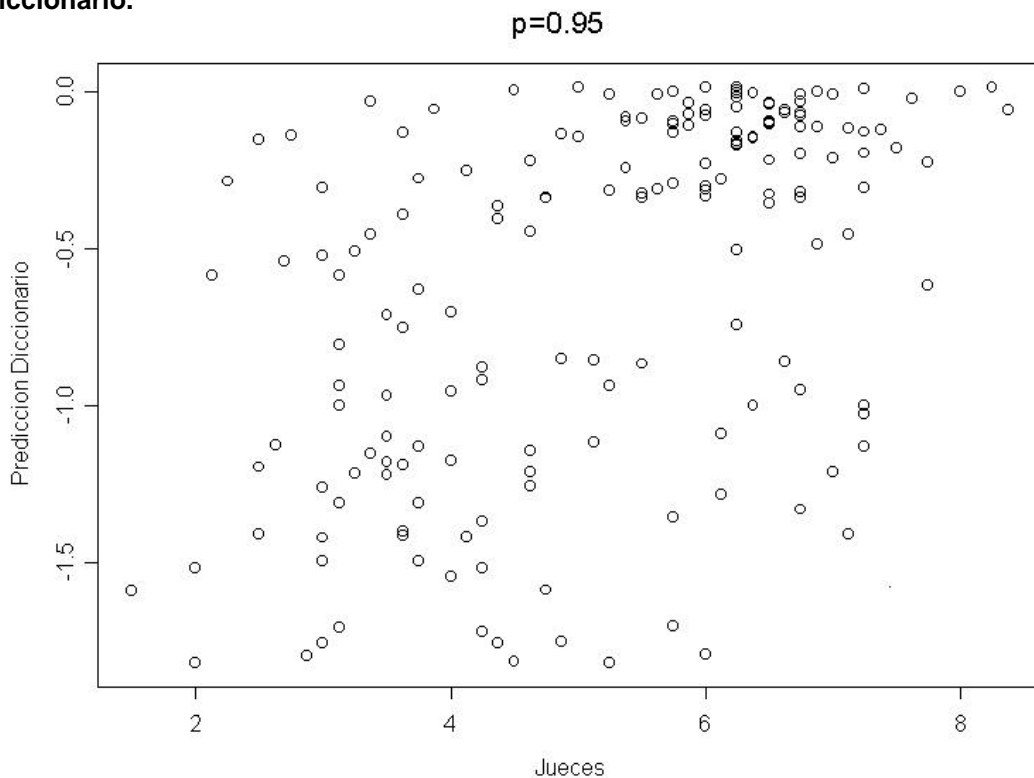
En este caso, existe una tendencia en los gráficos, sin embargo si se los observa en detalle, por ejemplo en el nivel cero de valor del diccionario, no se puede establecer ninguna calificación específica y única de los jueces. En ninguna de las gráficas hay una estructura definida, lo que estaría mostrando que a pesar del valor de la correlación obtenido el acuerdo entre ambos métodos no parece ser tan preciso.

Figura 9.- Gráfico de la estructura para el conjunto de textos de selección del parámetro de ponderación.



En este caso, a partir del gráfico se ve que no hay acuerdo entre los dos métodos.

Figura 10.- Gráfico de la estructura para el conjunto de textos no utilizados en la construcción del diccionario.



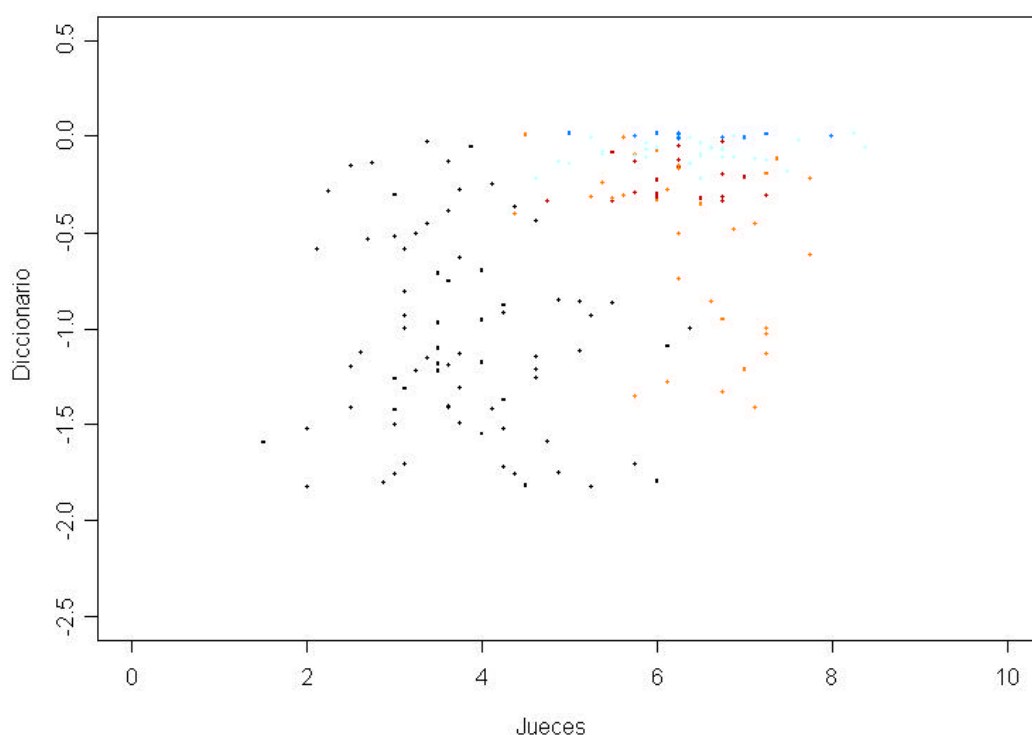
También en este caso se observa que a pesar del valor de la correlación obtenido no hay estructura y por lo tanto no hay buen acuerdo entre los métodos.

A partir de los gráficos antes presentados se podría concluir que no hay un buen nivel de acuerdo entre los dos métodos de análisis de la Actividad Referencial debido a que no se observa en ninguno de ellos una estructura definida.

Sin embargo, ¿cómo se explica la correlación obtenida?

La correlación reportada se produce debido a que en cada texto en particular no hay estructura ni correlación alta pero, en la medida que se agregan más textos, los puntos nuevos se incluyen en zonas donde no estaban los anteriores dando una correlación aparente. Este hecho puede observarse en la figura 11. Esto quiere decir, que cada texto en particular tiene correlaciones bajas y no muestra estructura ni acuerdo, y en la medida que se agregan más textos pueden ocurrir efectos compensatorios que resulten en una correlación más alta. Sin embargo, se sigue sin observar estructura. Esto explicaría también las correlaciones bajas obtenidas para los textos de prueba del parámetro de ponderación.

Figura 11.- Gráfico de puntos para la calificación de los jueces versus diccionario. Cada texto distinto se incluye en colores distintos.



Las posibles razones por las cuales no hay buen acuerdo son:

- a) Distintos errores metodológicos en la construcción del diccionario. Estos errores pueden ser:
 - i. El armado de las categorías. Tal como se explicó antes, no tiene sentido armar categorías que sólo sean utilizables para el diccionario que se construyó. Sería más adecuado crear categorías que fuesen utilizables en cualquier otro diccionario que se quisiera armar. De cualquier modo es también discutible el empleo de categorías ya que eso obliga a pasar de una escala de calificación ordinal como es la de los jueces, a una escala nominal lo que dificulta la comparación entre ambos métodos.
 - ii. La asignación arbitraria de la ponderación. El modo en que se asigna la ponderación como se mostró en las tablas 5 y 6 es arbitrario y no está debidamente explicitado.
 - iii. Las escalas. Relacionado con el punto anterior, el modo de asignación de la ponderación implica que el diccionario pasa de una escala de calificación de los jueces de 0-10 a una escala de -1 a 1. Esto no sería un problema si entre ellas se pudiese realizar un cambio de escalas, es decir, una transformación lineal entre ellas. Sin embargo, esto no es posible ya que el cero (valor medio en la escala del diccionario) no tiene relación con ningún otro valor de la escala de los jueces porque implica ausencia en el diccionario. Este hecho dificulta la comparación entre ambos métodos.
 - iv. Suma y promedio. Los resultados que se comparan por ambos métodos son el promedio de las cuatro escalas de los jueces para una unidad del texto y el valor de la suma o el promedio de la ponderación de

las palabras en esa unidad según el diccionario. Este procedimiento tiene varias falencias una de ellas es lo que sucede si hay repetición de una palabra, por ejemplo si el hablante dice: «no, no, no».

b) La pregunta sobre si ambos métodos están midiendo lo mismo.

Como conclusión general del trabajo, se confeccionó un diccionario en español que replica la metodología de elaboración utilizada para la generación de un diccionario para medir la Actividad Referencial en inglés. Los resultados obtenidos de la construcción son similares a los resultados reportados para el WRAD. Por lo tanto dicha creación bajo estos parámetros fue exitosa.

En este sentido, se desarrolló toda una herramienta computacional a través de distintos programas especialmente construidos. Dicho desarrollo resultó satisfactorio, ya que se obtuvo el diccionario SWRAD bajo los mismos parámetros que el WRAD.

Sin embargo, debido a un análisis detallado de la metodología utilizada, se puede concluir que el procedimiento empleado tiene aspectos metodológicos que pueden y deben ser optimizados. Al respecto, fue necesario realizar una amplia discusión acerca del uso de correlación y otros métodos estadísticos para la comparación de dos métodos. Se sistematizó así el análisis del acuerdo entre dos métodos y los primeros pasos necesarios para los procedimientos de modelización de estilos lingüísticos a través del uso de diccionarios. Esto no implica el descartar el diccionario sino el tener en cuenta la necesidad de desarrollar los métodos para su construcción antes de declararlo finalizado.

Finalmente, todo el esfuerzo en la construcción de diccionarios como este permite generar herramientas objetivas que den cuenta del accionar terapéutico. Es un esfuerzo que tiene sentido ya que establece otra mirada sobre el proceso terapéutico que permite mirar con detalle sus características y ampliar así el desarrollo de la práctica profesional. De los errores y de las críticas se aprenden y es a partir de estos que se construye en este proceso de ida y vuelta el conocimiento científico.

Referencias

- Altman, D. (1991). «Practical Statistics for Medical Research». Chapman & Hall. London, UK.
- Bucci, W. (1992). Scoring Referential Activity, Instruction for use Transcripts of Spoken Narrative and Texts, Ulm, Ulmer Textbank.
- Bucci, W. (1993). The development of emotional meaning in free association: A multiple code theory. En: A. Wilson y J. Gedo (compil.): Hierarchical Concepts in Psychoanalysis. New York, The Guilford Press.
- Bucci, W. (1997). Psychoanalysis and Cognitive Science: A multiple Code Theory. New York: Guilford Press.
- Bucci, W., Maskit, B. (2004). Building a Weighted Dictionary for Referential Activity. En Y. Qu, J.G. Shanahan; J. Wiebe (Eds.) *Exploring attitudes and affect in text; theories and applications*. N.Y. AAAI Press.
- Bucci, W. Adaptación de Avila, A., Vidal, J. «La puntuación de la Actividad Referencial». Versión preliminar. Unidad de Investigación en Psicología Clínica y Psicoterapia. Universidad de Salamanca.
- Dahl, H., Hölzer, M., Berry, J. (1992). «How to classify emotions for psychotherapy research». Ulmer Textbank, Universidad de Ulm, Alemania.
- Fertuck, E., Target, M., Mergenthaler, E., Clarkin, J. (2002). «Instrument development: from personality organization and temperament to psychotherapy process in individuals with BDP: Computerized reflective function: A psychotherapy process measure». Paper in panel. SPR 2002.
- Fonagy, P., Gergely, G., Jurist, E., Target, M. (2002). *Affect Regulation, Mentalization and the Development of the Self*. New York: Other Press.
- Maskit, B., Bucci, W., Roussos, A. «Capturing the flow of verbal interaction: the discourse attribute analysis program».
- Mergenthaler, E. (1996). «Emotion-Abstraction patterns in verbatim protocols: a new way of describing psychotherapeutic processes». *Journal of Consulting and Clinical Psychology*, 64, No. 6, 1306-1315.
- Mergenthaler, E., Bucci, W. (1999). «Linking verbal and non-verbal representations: Computer analysis of referential activity». *British Journal of Medical Psychology*, 72, 339-354.
- Roussos, A., Acosta, S., Juárez, C., Mergenthaler, E. (2001). «Introducción a las técnicas de investigación sobre procesos terapéuticos asistidos por computadora: Modelos de Ciclos Terapéuticos». *Interamerican Journal of Psychology*, Vol 35, Num 1, 53-77.
- Roussos, A., Leibovich, A. (2002). «La incidencia de la Actividad Referencial en el proceso de formulaciones clínicas. La importancia del marco teórico en la formulación de las inferencias clínicas en psicoterapia». *Intersubjetivo*, Madrid. V4 N1. p. 78-89
- Roussos A, Waizmann, V. Dorfman Lerner, B. y Polcaro, E. (2004). 35th Annual Meeting of the Society for Psychotherapy Research, Roma, Italia. «Types of Therapeutic Intervention in two Treatments in Relation to Weighted Referential Activity (WRAD)» en el simposio: Multi-Lingual Windows on the Referential Process; Application of New Referential Activity (RA) Dictionaries to Psychotherapy Transcripts in Italian, Spanish and Dutch. Moderator: Bucci, W., libro de resúmenes; pag. 210.
- Rosow, J. (1997). «The multiple code theory and the language of psychotherapy: the development of standardized measurement techniques». Disertación doctoral no publicada, Adelphi University, EEUU.

Apéndice I

Se transcribe en este apéndice el código fuente de los programas utilizados en la construcción del diccionario.

```
frecuency_function(texto, salida)
```

```
{
  z <- scan(texto, what = list(palabras = «»))
  words <- sort(z$palabras)
  k <- 1
  i <- 1
  j_2
  long <- length(words)
  while(i < long) {
    while(words[i] == words[j]){k_k+1
      j_j+1}
    if(words[i] != «/begin») {if(words[i] != «/end») {
      write(c(words[i], k), file = salida, ncol = 2, append = T)
      write(c(words[i], k), file = »frecu.txt«, ncol= 2, append = T)
    }}
    k <- 1
    print(i)
    i <- j
    j_j+1
  }
}
```

```
frecuencytotal_function(aaa)
```

```
{
  z <- scan(«frecu.txt», what = list(palabras = «»,frec=0))
  indice_sort.list(z$palabras)
  words <- sort(z$palabras)
  frecuencia_z$frec
  k <- 1
  i <- 1
  j_2
  long <- length(words)
  while(i < long) {k_frecuencia[indice[i]]
    while(words[i] == words[j]){k_k+frecuencia[indice[j]]
      j_j+1}
    if(words[i] != «/begin» && k >=aaa) {if(words[i] != «/end») {
      write(c(words[i], k), file = »palabras_diccionario.txt«, ncol = 2, append = T)
    }}
    print(i)
    i <- j
    j_j+1
  }
}
```

```
busqueda_function(entrada,salida,diccio)
```

```
{
  zz <- scan(diccio, what = list(palabras = «»,
    frecuencia = 0))
  words <- zz$palabras
  nw <- length(words)
  z <- scan(entrada, what = list(palabras = «»))
```

```

texto <- z$palabras
nt <- length(texto)
for(k in 1:nw) {
  i <- 1
  while(i <= nt) {
    b <- 0
    a <- 0
    if(texto[i] == «/begin») {
      a <- mean(as.numeric(c(texto[(i + 1):(i + 4)])))
      i <- i + 5
    }
    while(texto[i] != «/end») {
      if(words[k] == texto[i]) {
        b <- b + 1
      }
      i <- i + 1
    }
    i <- i + 1
    write(c(words[k], a, b), file = salida, ncol= 3, append = T)
    write(c(words[k], a, b), file = «S_begin_completo.txt», ncol= 3, append = T)
  }
}
}

```

armado de frecuencia de cada palabra dentro de la categoría
armado_function(nada) # la entrada es la salida2 de búsqueda

```

{
  z <- scan(«S_begin_completo.txt», what = list(palabras = «»,juez=0,frec=0))
  indice_sort.list(z$palabras)
  words <- sort(z$palabras)
  jueces_z$juez
  frecuencia_z$frec
  frecjuez_0*vector(length=6)
  num_0
  k <- 1
  i <- 1
  j_2
  long <- length(words)
  while(i < long) {frecjuez_c(0,0,0,0,0,0)
    num_seleccion(jueces[indice[i]]) # esto es para ver a que categoria entra
    frecjuez[num]_frecuencia[indice[i]]
    while(words[i] == words[j]){num_seleccion(jueces[indice[j]])
      frecjuez[num]_frecuencia[indice[j]]+frecjuez[num]
      j_j+1}

    write(c(words[i], frecjuez,sum(frecjuez)), file =»casi_puntaje.txt«, ncol = 2, append = T)

    print(i)
    i <- j
    j_j+1
  }
}

```

```

seleccion_function(a)
{
  if (a<2.75){b_1}
  if (2.875<=a<3.75){b_2}
  if (3.875<=a<4.75){b_3}

```

```

if (4.875<=a<5.75){b_4}
if (5.875<=a<7.125){b_5}
if (7.25<=a<=10){b_6}
return(b)}

```

```

puntaje_function(entrada,p) # la entrada es «casi_puntaje.txt»
{z <- scan(entrada, what = list(palabras = «»,a1=0,a2=0,a3=0,a4=0,a5=0,a6=0,suma=0))
words_z$palabras
frec_cbind(z$a1/sum(z$a1),z$a2/sum(z$a2),z$a3/sum(z$a3),z$a4/sum(z$a4),z$a5/sum(z$a5),z$a6/
sum(z$a6))# sobre la suma por columna
long_length(words)
aux_vector(length=6)
pesos_c(-1,-0.625,-.25,0.25,0.625,1)
#tres cosec es la del medio
#2 separadas eliminada NA
for(i in 1:long)
{aux_rank(frec[i,])
a_p*max(frec[i,])
sort.list(x)
for(j in 1:6){if(aux[j]!=6){}
x
[1] 2 5 1
> sort.list(x)
[1] 3 1 2
> order(x)
[1] 3 1 2
> rank(x)
[1] 2 3 1
>

```

```

find_function(entrada,salida,diccio)
{
zz <- scan(diccio, what = list(palabras = «», frecuencia = 0))
words <- zz$palabras
nw <- length(words)
z <- scan(entrada, what = list(palabras = «»))
texto <- z$palabras
nt <- length(texto)
auxiliar_c(1:nt)
beg_texto==»/begin»
end_texto==»/end»
tot_end%*%end
indice_sort.list(end)
final_indice[(nt-tot+1):nt]
indice_sort.list(beg)
comienzo_indice[(nt-tot+1):nt]
for(k in 1:nw) {
cont_texto==words[k]
for(i in 1:tot){
b_cont[comienzo[i]:final[i]]%*%cont[comienzo[i]:final[i]]
a <- mean(as.numeric(c(texto[(comienzo[i] + 1):(comienzo[i] + 4)])))
write(c(words[k], a, b), file = salida, ncol= 3, append = T)
write(c(words[k], a, b), file = «prueba.txt», ncol= 3, append = T)
#»S_begin_completo.txt», ncol= 3, append = T)
}
}}

```

```

puntaje_function(entrada,salida, p)
{
  z <- scan(entrada, what = list(palabras = «», a1 = 0, a2 = 0,
    a3 = 0, a4 = 0, a5 = 0, a6 = 0, suma = 0))
  words <- z$palabras
  frec <- cbind(z$a1/sum(z$a1), z$a2/sum(z$a2), z$a3/sum(z$a3), z$a
    a4/sum(z$a4), z$a5/sum(z$a5), z$a6/sum(z$a6))
  long <- length(words)
  for(i in 1:long) {
    write(c(z$palabras[i], frec[i, ]), file =
      «Tabla_1.txt», ncol = 7, append = T)
  }
  aux <- vector(length = 6)
  pesos <- c(-1, -0.625, -0.25, 0.25, 0.625, 1)
  for(i in 1:long) {
    sup <- max(frec[i, ])
    sup2 <- p * sup
    ss <- 0
    if(sup == frec[i, 1]) {
      if(sup2 > frec[i, 2] && sup2 > frec[i, 3] &&
        sup2 > frec[i, 4] && sup2 > frec[i, 5] &&
        sup2 > frec[i, 6]) {
        write(c(z$palabras[i], pesos[1]), file
          = salida, ncol = 2, append =
            T)
        ss <- 1
      }
      if(sup2 < frec[i, 2] && sup2 > frec[i, 3] &&
        sup2 > frec[i, 4] && sup2 > frec[i, 5] &&
        sup2 > frec[i, 6]) {
        write(c(z$palabras[i], -0.875), file =
          salida, ncol = 2, append = T)
        ss <- 1
      }
      if(sup2 < frec[i, 2] && sup2 < frec[i, 3] &&
        sup2 > frec[i, 4] && sup2 > frec[i, 5] &&
        sup2 > frec[i, 6]) {
        write(c(z$palabras[i], pesos[2]), file
          = salida, ncol = 2, append =
            T)
        ss <- 1
      }
      if(ss == 0) {
        write(c(z$palabras[i], «NA»), file =
          salida, ncol = 2, append = T)
        ss <- 1
      }
    }
  }

  if(sup == frec[i, 2] && ss == 0) {
    if(sup2 > frec[i, 1] && sup2 > frec[i, 3] &&
      sup2 > frec[i, 4] && sup2 > frec[i, 5] &&
      sup2 > frec[i, 6]) {
      write(c(z$palabras[i], pesos[2]), file
        = salida, ncol = 2, append =
          T)
      ss <- 1
    }
  }
}

```

```

}
if(sup2 < frec[i, 1] && sup2 > frec[i, 3] &&
  sup2 > frec[i, 4] && sup2 > frec[i, 5] &&
  sup2 > frec[i, 6]) {
  write(c(z$palabras[i], -0.75), file =
    salida, ncol = 2, append = T)
  ss <- 1
}
if(sup2 > frec[i, 1] && sup2 < frec[i, 3] &&
  sup2 > frec[i, 4] && sup2 > frec[i, 5] &&
  sup2 > frec[i, 6]) {
  write(c(z$palabras[i], -0.5), file =
    salida, ncol = 2, append = T)
  ss <- 1
}
if(sup2 < frec[i, 1] && sup2 < frec[i, 3] &&
  sup2 > frec[i, 4] && sup2 > frec[i, 5] &&
  sup2 > frec[i, 6]) {
  write(c(z$palabras[i], pesos[2]), file
    = salida, ncol = 2, append =
    T)
  ss <- 1
}
if(sup2 > frec[i, 1] && sup2 < frec[i, 3] &&
  sup2 < frec[i, 4] && sup2 > frec[i, 5] &&
  sup2 > frec[i, 6]) {
  write(c(z$palabras[i], pesos[3]), file
    = salida, ncol = 2, append =
    T)
  ss <- 1
}
}
if(ss == 0) {
  write(c(z$palabras[i], «NA»), file =
    salida, ncol = 2, append = T)
  ss <- 1
}
}

```

```

if(sup == frec[i, 3] && ss == 0) {
  if(sup2 > frec[i, 1] && sup2 > frec[i, 2] &&
    sup2 > frec[i, 4] && sup2 > frec[i, 5] &&
    sup2 > frec[i, 6]) {
    write(c(z$palabras[i], pesos[3]), file
      = salida, ncol = 2, append =
      T)
    ss <- 1
  }
  if(sup2 > frec[i, 1] && sup2 < frec[i, 2] &&
    sup2 > frec[i, 4] && sup2 > frec[i, 5] &&
    sup2 > frec[i, 6]) {
    write(c(z$palabras[i], -0.375), file =
      salida, ncol = 2, append = T)
    ss <- 1
  }
  if(sup2 > frec[i, 1] && sup2 < frec[i, 4] &&
    sup2 > frec[i, 2] && sup2 > frec[i, 5] &&
    sup2 > frec[i, 6]) {

```

```

write(c(z$palabras[i], -0.125), file =
  salida, ncol = 2, append = T)
ss <- 1
}
if(sup2 > frec[i, 1] && sup2 < frec[i, 2] &&
  sup2 < frec[i, 4] && sup2 > frec[i, 5] &&
  sup2 > frec[i, 6]) {
write(c(z$palabras[i], pesos[3]), file
  = salida, ncol = 2, append =
  T)
ss <- 1
}
if(sup2 > frec[i, 1] && sup2 > frec[i, 2] &&
  sup2 < frec[i, 4] && sup2 < frec[i, 5] &&
  sup2 > frec[i, 6]) {
write(c(z$palabras[i], pesos[4]), file
  = salida, ncol = 2, append =
  T)
ss <- 1
}
if(ss == 0) {
write(c(z$palabras[i], «NA»), file =
  salida, ncol = 2, append = T)
ss <- 1
}
}

```

```

if(sup == frec[i, 4] && ss == 0) {
if(sup2 > frec[i, 1] && sup2 > frec[i, 2] &&
  sup2 > frec[i, 3] && sup2 > frec[i, 5] &&
  sup2 > frec[i, 6]) {
write(c(z$palabras[i], pesos[4]), file
  = salida, ncol = 2, append =
  T)
ss <- 1
}
if(sup2 > frec[i, 1] && sup2 > frec[i, 2] &&
  sup2 < frec[i, 3] && sup2 > frec[i, 5] &&
  sup2 > frec[i, 6]) {
write(c(z$palabras[i], 0.125), file =
  salida, ncol = 2, append = T)
ss <- 1
}
if(sup2 > frec[i, 1] && sup2 > frec[i, 2] &&
  sup2 > frec[i, 3] && sup2 < frec[i, 5] &&
  sup2 > frec[i, 6]) {
write(c(z$palabras[i], 0.375), file =
  salida, ncol = 2, append = T)
ss <- 1
}
if(sup2 > frec[i, 1] && sup2 > frec[i, 2] &&
  sup2 < frec[i, 3] && sup2 < frec[i, 5] &&
  sup2 > frec[i, 6]) {
write(c(z$palabras[i], pesos[4]), file
  = salida, ncol = 2, append =
  T)
ss <- 1
}
}

```

```

if(sup2 > frec[i, 1] && sup2 > frec[i, 2] &&
  sup2 > frec[i, 3] && sup2 < frec[i, 5] &&
  sup2 < frec[i, 6]) {
  write(c(z$palabras[i], pesos[5]), file
    = salida, ncol = 2, append =
    T)
  ss <- 1
}
if(ss == 0) {
  write(c(z$palabras[i], «NA»), file =
    salida, ncol = 2, append = T)
  ss <- 1
}
}

```

```

if(sup == frec[i, 5] && ss == 0) {
  if(sup2 > frec[i, 1] && sup2 > frec[i, 2] &&
    sup2 > frec[i, 3] && sup2 > frec[i, 4] &&
    sup2 > frec[i, 6]) {
    write(c(z$palabras[i], pesos[5]), file
      = salida, ncol = 2, append =
      T)
    ss <- 1
  }
  if(sup2 > frec[i, 1] && sup2 > frec[i, 2] &&
    sup2 > frec[i, 3] && sup2 < frec[i, 4] &&
    sup2 > frec[i, 6]) {
    write(c(z$palabras[i], 0.5), file =
      salida, ncol = 2, append = T)
    ss <- 1
  }
  if(sup2 > frec[i, 1] && sup2 > frec[i, 2] &&
    sup2 > frec[i, 3] && sup2 > frec[i, 4] &&
    sup2 < frec[i, 6]) {
    write(c(z$palabras[i], 0.75), file =
      salida, ncol = 2, append = T)
    ss <- 1
  }
  if(sup2 > frec[i, 1] && sup2 > frec[i, 2] &&
    sup2 > frec[i, 3] && sup2 < frec[i, 4] &&
    sup2 < frec[i, 6]) {
    write(c(z$palabras[i], pesos[5]), file
      = salida, ncol = 2, append =
      T)
    ss <- 1
  }
  if(ss == 0) {
    write(c(z$palabras[i], «NA»), file =
      salida, ncol = 2, append = T)
    ss <- 1
  }
}
}

```

```

if(sup == frec[i, 6] && ss == 0) {
  if(sup2 > frec[i, 1] && sup2 > frec[i, 2] &&
    sup2 > frec[i, 3] && sup2 > frec[i, 4] &&
    sup2 > frec[i, 5]) {
    write(c(z$palabras[i], pesos[6]), file

```

```

        = salida, ncol = 2, append =
        T)
        ss <- 1
    }
    if(sup2 > frec[i, 1] && sup2 > frec[i, 2] &&
        sup2 > frec[i, 3] && sup2 > frec[i, 4] &&
        sup2 < frec[i, 5]) {
        write(c(z$palabras[i], 0.875), file =
            salida, ncol = 2, append = T)
        ss <- 1
    }
    if(ss == 0) {
        write(c(z$palabras[i], «NA»), file =
            salida, ncol = 2, append = T)
        ss <- 1
    }
}
}
}

```

```

jueces_function(entrada, comienfin)
{
  a <- vector(length = 4)
  z <- scan(entrada, what = list(palabras = «»))
  texto <- z$palabras
  nt <- length(texto)
  i <- 1
  j_0
  while(i <= nt) {
    if(texto[i] == «/begin») {
      a <- as.numeric(c(texto[(i + 1):(i + 4)]))
      i <- i + 5
      j_i
      write(a, file = «jueces_todo.txt», ncol = 4,
          append = T)
      write(mean(a), file = «jueces_media.txt», ncol
          = 1, append = T)
      write(var(a), file = «jueces_varianza.txt»,
          ncol = 1, append = T)
    }
    if(texto[i] == «/end») {
      write((i-j), file = «jueces_length.txt», ncol = 1, append = T)
      write(c(j,(i-1)), file = «jueces_comien_fin.txt», ncol = 2, append = T)
      write(c(j,(i-1)), file = comienfin, ncol = 2, append = T)
      i <- i + 1
    }
  }
}

```

```

analisis_function(entrada,juecescomienfin,diccio)
{
  suma_0
  n_0
  zz <- scan(diccio, what = list(palabras = «», valor = 0))
  puntaje_zz$valor
  words <- zz$palabras
  nw <- length(words)
  zz <- scan(juecescomienfin, what = list(start=0,end=0))

```



```

z <- scan(entrada, what = list(palabras = «»))
texto <- z$palabras
nt <- length(texto)
cont_texto==»/begin»
nparrafos_cont%*%cont
for(i in 1:nparrafos) {text_texto[zz$start[i]:zz$end[i]]
  n_0
for(k in 1:nw) {
cont_text==words[k]
b_cont%*%cont
if(puntaje[k]!=»NA»){suma_b*puntaje[k]+suma
  n_n+1}
}
write(suma/n, file = «promedio_final.txt», ncol= 1, append = T)
write(suma, file = «suma_final.txt», ncol= 1, append = T)
}
}

```

```

esto es para textos nuevos
analisisnew_function(entrada,diccio,promediofinal,sumafinal)
{
  suma_0
  n_0
  zz <- scan(diccio, what = list(palabras = «», valor = 0))
  puntaje_zz$valor
  words <- zz$palabras
  juez_0
  nw <- length(words)
  z <- scan(entrada, what = list(palabras = «»))
  texto <- z$palabras
  nt <- length(texto)
  cont_texto==»/begin»
  cont2_texto==»/end»
  nparrafos_cont%*%cont
  start_vector(length=nparrafos)
  end_vector(length=nparrafos)
  j_1
  jj_1
  for (i in 1:nt){if(cont[i]==T){start[jj]_i+5
    j_j+1}
  if(cont2[i]==T){end[jj]_(i-1)
    jj_jj+1}
  }
  for(i in 1:nparrafos) {text_texto[start[i]:end[i]]
  n_0
for(k in 1:nw) {
cont_text==words[k]
b_cont%*%cont
if(puntaje[k]!=»NA»){suma_b*puntaje[k]+suma
  n_n+1}
}
juez <- mean(as.numeric(c(texto[(start[i]-4):(start[i]-1)])))
write(c(juez,suma/n), file =promediofinal, ncol= 2, append = T)
write(c(juez,suma), file =sumafinal, ncol= 2, append = T)
}
}

```

Apéndice 2

Diccionario completo con los puntajes asociados ordenados por frecuencia.

Palabra diccionario	Categoría 1	Categoría 2	Categoría 3	Categoría 4	Categoría 5	Categoría 6	Frecuencia total	Ponderación P=0.95
que	0.0705	0.0665	0.0729	0.0704	0.0640	0.0641	3374	NA
no	0.0493	0.0627	0.0340	0.0297	0.0401	0.0393	2157	-0.625
de	0.0322	0.0345	0.0467	0.0395	0.0437	0.0398	2005	-0.25
y	0.0331	0.0368	0.0345	0.0437	0.0400	0.0432	2002	NA
a	0.0251	0.0271	0.0323	0.0313	0.0302	0.0297	1477	-0.125
la	0.0229	0.0225	0.0326	0.0325	0.0276	0.0325	1430	NA
en	0.0210	0.0209	0.0280	0.0258	0.0223	0.0250	1187	-0.25
me	0.0219	0.0241	0.0207	0.0213	0.0236	0.0253	1168	NA
el	0.0185	0.0200	0.0248	0.0163	0.0217	0.0245	1078	NA
es	0.0295	0.0248	0.0195	0.0146	0.0170	0.0178	1006	-1
lo	0.0239	0.0153	0.0165	0.0165	0.0198	0.0177	918	-1
yo	0.0212	0.0171	0.0158	0.0185	0.0166	0.0171	877	-1
un	0.0114	0.0108	0.0165	0.0160	0.0145	0.0157	714	-0.125
se	0.0122	0.0108	0.0089	0.0139	0.0153	0.0172	683	1
con	0.0093	0.0147	0.0159	0.0146	0.0119	0.0139	670	-0.25
como	0.0168	0.0176	0.0121	0.0116	0.0109	0.0121	664	-0.75
por	0.0115	0.0112	0.0116	0.0130	0.0107	0.0163	626	1
una	0.0105	0.0115	0.0106	0.0126	0.0132	0.0129	610	0.625
mi	0.0059	0.0094	0.0104	0.0167	0.0137	0.0117	600	0.25
si	0.0117	0.0113	0.0144	0.0087	0.0124	0.0110	587	-0.25
pero	0.0131	0.0122	0.0098	0.0100	0.0103	0.0142	578	1
porque	0.0156	0.0105	0.0100	0.0092	0.0095	0.0116	546	-1
o	0.0139	0.0118	0.0087	0.0087	0.0087	0.0122	533	-1
los	0.0047	0.0050	0.0104	0.0162	0.0117	0.0117	506	0.25
más	0.0070	0.0072	0.0104	0.0108	0.0076	0.0042	375	0.125
las	0.0054	0.0065	0.0098	0.0078	0.0073	0.0076	372	-0.25
muy	0.0049	0.0065	0.0087	0.0096	0.0083	0.0061	369	0.25
bueno	0.0102	0.0089	0.0078	0.0059	0.0069	0.0047	360	-1
te	0.0129	0.0110	0.0086	0.0046	0.0037	0.0055	359	-1
para	0.0047	0.0073	0.0069	0.0066	0.0070	0.0076	345	NA
era	0.0024	0.0041	0.0055	0.0066	0.0097	0.0073	326	0.625
sí	0.0114	0.0104	0.0038	0.0032	0.0042	0.0038	294	-1
eso	0.0059	0.0057	0.0063	0.0078	0.0056	0.0041	286	0.25
le	0.0044	0.0027	0.0028	0.0060	0.0066	0.0089	281	1
cuando	0.0036	0.0037	0.0035	0.0076	0.0063	0.0063	267	0.25
qué	0.0061	0.0059	0.0049	0.0034	0.0049	0.0059	265	NA
sé	0.0086	0.0098	0.0018	0.0014	0.0039	0.0051	261	-0.625
entonces	0.0059	0.0054	0.0083	0.0068	0.0046	0.0023	258	-0.25
cosas	0.0044	0.0069	0.0034	0.0037	0.0044	0.0057	246	-0.625
todo	0.0058	0.0043	0.0046	0.0050	0.0047	0.0052	246	-1
sea	0.0075	0.0056	0.0041	0.0027	0.0031	0.0040	217	-1
mucho	0.0037	0.0051	0.0023	0.0059	0.0044	0.0030	203	0.25
fue	0.0027	0.0030	0.0026	0.0062	0.0049	0.0040	200	0.25
su	0.0051	0.0026	0.0017	0.0032	0.0055	0.0040	195	0.625
del	0.0024	0.0038	0.0047	0.0027	0.0044	0.0041	194	-0.25
algo	0.0022	0.0031	0.0023	0.0027	0.0036	0.0065	186	1
ella	0.0019	0.0031	0.0025	0.0039	0.0056	0.0031	181	0.625
al	0.0029	0.0017	0.0044	0.0041	0.0039	0.0036	174	-0.25
uno	0.0022	0.0031	0.0049	0.0030	0.0040	0.0031	174	-0.25
está	0.0024	0.0035	0.0047	0.0020	0.0041	0.0030	171	-0.25
tu	0.0029	0.0048	0.0057	0.0048	0.0032	0.0003	166	-0.25
ya	0.0041	0.0037	0.0034	0.0025	0.0036	0.0026	166	-1
eh	0.0058	0.0064	0.0018	0.0034	0.0015	0.0024	165	-0.625
casa	0.0020	0.0013	0.0006	0.0016	0.0052	0.0057	164	1
siempre	0.0039	0.0036	0.0017	0.0044	0.0026	0.0034	160	0.25
ahora	0.0032	0.0031	0.0037	0.0027	0.0030	0.0033	159	-0.25
ahí	0.0025	0.0040	0.0026	0.0023	0.0030	0.0038	159	NA
nada	0.0056	0.0044	0.0018	0.0018	0.0039	0.0015	159	-1
tiene	0.0027	0.0040	0.0025	0.0039	0.0039	0.0018	158	NA
después	0.0008	0.0024	0.0020	0.0041	0.0048	0.0030	156	0.625
él	0.0032	0.0026	0.0020	0.0025	0.0043	0.0027	153	0.625
también	0.0029	0.0035	0.0026	0.0028	0.0031	0.0029	151	-0.625
bien	0.0039	0.0038	0.0031	0.0025	0.0025	0.0023	148	-0.875
así	0.0029	0.0034	0.0025	0.0041	0.0019	0.0034	146	0.25
este	0.0025	0.0029	0.0035	0.0048	0.0025	0.0021	145	0.25

creo	0.0036	0.0019	0.0066	0.0034	0.0020	0.0017	143	-0.25
había	0.0015	0.0010	0.0031	0.0046	0.0034	0.0031	141	0.25
tengo	0.0034	0.0026	0.0012	0.0037	0.0021	0.0036	136	NA
claro	0.0034	0.0027	0.0021	0.0023	0.0029	0.0025	134	-1
años	0.0007	0.0010	0.0023	0.0037	0.0036	0.0035	133	0.375
hay	0.0027	0.0033	0.0026	0.0025	0.0022	0.0027	133	-0.625
cosa	0.0025	0.0036	0.0021	0.0032	0.0022	0.0022	129	-0.625
tenía	0.0015	0.0016	0.0023	0.0027	0.0035	0.0029	129	0.625
esto	0.0029	0.0044	0.0023	0.0028	0.0021	0.0011	125	-0.625
estaba	0.0014	0.0021	0.0012	0.0028	0.0029	0.0032	122	1
cómo	0.0024	0.0026	0.0040	0.0037	0.0015	0.0018	121	-0.25
usted	0.0008	0.0010	0.0017	0.0021	0.0027	0.0046	121	1
mamá	0.0012	0.0020	0.0011	0.0025	0.0026	0.0036	118	1
nos	0.0019	0.0017	0.0012	0.0023	0.0036	0.0021	115	0.625
otro	0.0025	0.0015	0.0025	0.0034	0.0021	0.0023	115	0.25
vez	0.0020	0.0024	0.0020	0.0021	0.0024	0.0024	115	NA
dos	0.0019	0.0009	0.0025	0.0034	0.0016	0.0036	114	1
hacer	0.0012	0.0026	0.0021	0.0025	0.0027	0.0020	113	0.625
ver	0.0037	0.0031	0.0018	0.0021	0.0019	0.0015	113	-1
papá	0.0003	0.0010	0.0012	0.0014	0.0026	0.0046	110	1
ese	0.0012	0.0015	0.0026	0.0028	0.0032	0.0014	109	0.625
va	0.0034	0.0038	0.0021	0.0009	0.0014	0.0015	106	-0.625
pues	0.0076	0.0028	0.0032	0.0014	0.0005	0.0000	105	-1
todos	0.0022	0.0019	0.0014	0.0016	0.0025	0.0022	104	0.625
puede	0.0025	0.0022	0.0011	0.0014	0.0025	0.0020	103	NA
digo	0.0036	0.0033	0.0011	0.0014	0.0017	0.0014	101	-1
son	0.0007	0.0023	0.0017	0.0020	0.0032	0.0011	100	0.625
padre	0.0017	0.0012	0.0006	0.0027	0.0029	0.0021	99	0.625
vos	0.0053	0.0043	0.0008	0.0007	0.0006	0.0013	99	-1
madre	0.0034	0.0019	0.0006	0.0020	0.0030	0.0007	97	-1
ser	0.0036	0.0024	0.0011	0.0014	0.0019	0.0014	96	-1
ni	0.0037	0.0015	0.0015	0.0004	0.0027	0.0011	94	-1
ha	0.0027	0.0019	0.0046	0.0016	0.0008	0.0009	92	-0.25
dice	0.0014	0.0010	0.0015	0.0011	0.0018	0.0033	91	1
otra	0.0027	0.0014	0.0020	0.0012	0.0024	0.0011	91	-1
familia	0.0014	0.0008	0.0032	0.0023	0.0019	0.0014	89	-0.25
ejemplo	0.0005	0.0015	0.0038	0.0028	0.0008	0.0018	86	-0.25
hace	0.0015	0.0021	0.0020	0.0018	0.0021	0.0008	86	NA
menos	0.0027	0.0015	0.0012	0.0014	0.0024	0.0009	86	-1
poco	0.0036	0.0016	0.0025	0.0011	0.0015	0.0009	86	-1
día	0.0017	0.0009	0.0006	0.0018	0.0019	0.0027	85	1
esa	0.0017	0.0010	0.0008	0.0027	0.0020	0.0019	85	0.25
voy	0.0032	0.0024	0.0011	0.0014	0.0013	0.0012	85	-1
quiero	0.0003	0.0013	0.0017	0.0020	0.0012	0.0031	83	1
trabajo	0.0017	0.0016	0.0026	0.0020	0.0009	0.0018	83	-0.25
momento	0.0005	0.0007	0.0032	0.0027	0.0017	0.0014	82	-0.25
ir	0.0010	0.0009	0.0020	0.0009	0.0018	0.0022	79	1
estoy	0.0031	0.0029	0.0020	0.0014	0.0005	0.0007	78	-0.875
recuerdo	0.0002	0.0015	0.0012	0.0034	0.0015	0.0017	78	0.25
esta	0.0012	0.0017	0.0020	0.0016	0.0012	0.0014	75	-0.25
mismo	0.0014	0.0008	0.0023	0.0014	0.0016	0.0015	75	-0.25
dijo	0.0010	0.0019	0.0008	0.0018	0.0019	0.0012	74	NA
parece	0.0015	0.0014	0.0025	0.0012	0.0010	0.0016	74	-0.25
tiempo	0.0010	0.0016	0.0014	0.0011	0.0018	0.0013	72	0.625
veces	0.0008	0.0010	0.0023	0.0018	0.0020	0.0006	71	-0.25
tan	0.0017	0.0017	0.0014	0.0007	0.0005	0.0023	69	1
ah	0.0005	0.0013	0.0011	0.0016	0.0017	0.0015	68	0.625
decir	0.0017	0.0023	0.0009	0.0011	0.0008	0.0013	67	-0.625
estar	0.0020	0.0016	0.0008	0.0009	0.0013	0.0013	67	-1
hasta	0.0008	0.0013	0.0012	0.0009	0.0012	0.0021	67	1
mis	0.0008	0.0009	0.0012	0.0012	0.0020	0.0012	67	0.625
persona	0.0008	0.0008	0.0011	0.0021	0.0022	0.0007	67	0.5
mejor	0.0014	0.0005	0.0025	0.0023	0.0013	0.0007	66	-0.25
ellos	0.0005	0.0009	0.0026	0.0014	0.0011	0.0014	65	-0.25
dije	0.0019	0.0014	0.0012	0.0014	0.0008	0.0012	63	-1
nunca	0.0000	0.0020	0.0005	0.0009	0.0018	0.0014	63	-0.625
cada	0.0012	0.0007	0.0017	0.0005	0.0015	0.0014	62	-0.25
iba	0.0005	0.0017	0.0005	0.0018	0.0014	0.0012	62	NA
tal	0.0017	0.0023	0.0017	0.0005	0.0008	0.0006	61	-0.625
tema	0.0020	0.0012	0.0005	0.0016	0.0007	0.0017	61	-1
acuerdo	0.0002	0.0002	0.0005	0.0004	0.0019	0.0026	60	1
pasa	0.0020	0.0003	0.0003	0.0014	0.0015	0.0013	59	-1

nosotros	0.0008	0.0007	0.0006	0.0011	0.0018	0.0013	58	0.625
puedo	0.0014	0.0017	0.0005	0.0007	0.0009	0.0015	58	-0.625
antes	0.0014	0.0012	0.0006	0.0009	0.0016	0.0008	57	0.625
vida	0.0008	0.0016	0.0006	0.0023	0.0008	0.0010	57	0.25
mal	0.0020	0.0024	0.0000	0.0014	0.0006	0.0007	56	-0.625
problema	0.0020	0.0005	0.0003	0.0000	0.0012	0.0022	56	1
tipo	0.0008	0.0015	0.0017	0.0002	0.0012	0.0010	56	-0.25
realmente	0.0015	0.0008	0.0020	0.0009	0.0015	0.0002	55	-0.25
relación	0.0010	0.0010	0.0025	0.0009	0.0005	0.0011	55	-0.25
donde	0.0003	0.0009	0.0006	0.0016	0.0015	0.0010	54	0.375
fui	0.0015	0.0014	0.0006	0.0016	0.0008	0.0009	54	NA
alguna	0.0015	0.0005	0.0005	0.0007	0.0009	0.0020	53	1
gente	0.0008	0.0012	0.0011	0.0023	0.0006	0.0009	53	0.25
manera	0.0020	0.0020	0.0012	0.0005	0.0005	0.0005	52	-0.875
soy	0.0029	0.0020	0.0005	0.0004	0.0005	0.0007	52	-1
sus	0.0019	0.0012	0.0005	0.0012	0.0011	0.0007	52	-1
importante	0.0003	0.0005	0.0006	0.0012	0.0015	0.0013	50	0.625
mí	0.0007	0.0007	0.0008	0.0011	0.0009	0.0016	50	1
año	0.0002	0.0010	0.0005	0.0004	0.0018	0.0010	49	0.625
cuenta	0.0007	0.0008	0.0008	0.0011	0.0010	0.0013	49	1
trabajar	0.0002	0.0006	0.0023	0.0011	0.0014	0.0004	49	-0.25
tres	0.0000	0.0005	0.0023	0.0018	0.0010	0.0007	49	1
haber	0.0015	0.0005	0.0000	0.0012	0.0009	0.0015	48	NA
sido	0.0012	0.0006	0.0009	0.0005	0.0007	0.0017	48	1
lado	0.0008	0.0009	0.0002	0.0018	0.0010	0.0009	47	0.25
tener	0.0012	0.0015	0.0008	0.0009	0.0008	0.0006	47	-0.625
verdad	0.0025	0.0013	0.0009	0.0009	0.0005	0.0003	47	-1
hijos	0.0005	0.0006	0.0002	0.0007	0.0012	0.0017	46	1
sabes	0.0027	0.0016	0.0018	0.0005	0.0000	0.0001	46	-1
decía	0.0003	0.0005	0.0012	0.0012	0.0013	0.0007	45	0.625
desde	0.0010	0.0008	0.0008	0.0011	0.0008	0.0010	45	NA
hecho	0.0010	0.0007	0.0009	0.0007	0.0009	0.0009	44	-1
sin	0.0012	0.0007	0.0008	0.0007	0.0008	0.0010	44	-1
acá	0.0007	0.0010	0.0006	0.0000	0.0009	0.0013	43	1
parte	0.0020	0.0003	0.0006	0.0009	0.0012	0.0003	43	-1
analista	0.0000	0.0014	0.0028	0.0009	0.0004	0.0002	42	-0.25
cuatro	0.0002	0.0000	0.0014	0.0016	0.0009	0.0010	42	1
poder	0.0003	0.0008	0.0020	0.0007	0.0005	0.0008	42	-0.25
entre	0.0010	0.0007	0.0008	0.0005	0.0007	0.0011	41	1
fuera	0.0002	0.0001	0.0009	0.0002	0.0013	0.0014	41	1
institución	0.0007	0.0002	0.0012	0.0016	0.0014	0.0000	41	0.25
personas	0.0000	0.0002	0.0011	0.0016	0.0013	0.0006	41	0.25
algún	0.0002	0.0002	0.0018	0.0004	0.0011	0.0008	40	-0.25
análisis	0.0000	0.0003	0.0041	0.0007	0.0004	0.0001	40	-0.25
bastante	0.0007	0.0014	0.0005	0.0004	0.0010	0.0006	40	-0.625
sentido	0.0014	0.0006	0.0017	0.0007	0.0008	0.0001	40	-0.25
todas	0.0003	0.0005	0.0012	0.0007	0.0012	0.0007	40	-0.25
tú	0.0020	0.0014	0.0014	0.0002	0.0004	0.0001	40	-1
caso	0.0008	0.0007	0.0012	0.0002	0.0008	0.0007	39	-0.25
sino	0.0008	0.0005	0.0009	0.0005	0.0009	0.0008	39	NA
días	0.0012	0.0009	0.0008	0.0004	0.0008	0.0005	38	-1
eran	0.0007	0.0003	0.0015	0.0007	0.0005	0.0010	38	-0.25
formación	0.0000	0.0005	0.0035	0.0007	0.0005	0.0000	38	-0.25
muchas	0.0008	0.0008	0.0006	0.0009	0.0011	0.0003	38	0.625
otras	0.0005	0.0010	0.0003	0.0012	0.0007	0.0007	38	0.25
siento	0.0008	0.0015	0.0008	0.0004	0.0006	0.0005	38	-0.625
están	0.0007	0.0008	0.0012	0.0011	0.0003	0.0007	37	-0.25
forma	0.0002	0.0008	0.0011	0.0009	0.0008	0.0007	37	-0.25
toda	0.0005	0.0003	0.0003	0.0011	0.0010	0.0009	37	0.25
tus	0.0007	0.0015	0.0020	0.0002	0.0001	0.0005	37	-0.25
además	0.0015	0.0003	0.0003	0.0004	0.0006	0.0011	36	-1
he	0.0012	0.0007	0.0011	0.0007	0.0007	0.0003	36	-1
vamos	0.0008	0.0009	0.0009	0.0005	0.0008	0.0004	36	-0.5
grande	0.0003	0.0000	0.0006	0.0002	0.0008	0.0016	35	1
han	0.0005	0.0005	0.0017	0.0011	0.0008	0.0001	35	-0.25
hoy	0.0005	0.0002	0.0006	0.0014	0.0005	0.0011	35	0.25
mujer	0.0002	0.0005	0.0002	0.0004	0.0019	0.0003	35	0.625
veo	0.0017	0.0013	0.0002	0.0011	0.0005	0.0001	35	-1
auto	0.0000	0.0001	0.0000	0.0002	0.0022	0.0004	34	0.625
da	0.0017	0.0003	0.0006	0.0009	0.0003	0.0007	34	-1
dígame	0.0025	0.0006	0.0002	0.0004	0.0006	0.0003	34	-1
les	0.0008	0.0003	0.0014	0.0007	0.0006	0.0005	34	-0.25

nj	0.0025	0.0022	0.0000	0.0000	0.0000	0.0000	34	-1
vas	0.0012	0.0013	0.0003	0.0000	0.0007	0.0005	34	-0.625
hermana	0.0002	0.0000	0.0008	0.0012	0.0012	0.0004	33	0.375
mas	0.0003	0.0005	0.0000	0.0007	0.0007	0.0013	33	1
pensar	0.0005	0.0014	0.0009	0.0004	0.0005	0.0004	33	-0.625
quiere	0.0003	0.0007	0.0006	0.0005	0.0006	0.0009	33	1
tienen	0.0003	0.0007	0.0005	0.0005	0.0007	0.0009	33	1
aquí	0.0005	0.0009	0.0015	0.0009	0.0002	0.0003	32	-0.25
doy	0.0007	0.0005	0.0003	0.0005	0.0002	0.0016	32	1
fueron	0.0000	0.0001	0.0018	0.0007	0.0006	0.0007	32	-0.25
gusta	0.0002	0.0003	0.0008	0.0011	0.0006	0.0008	32	0.25
hemos	0.0022	0.0014	0.0005	0.0002	0.0001	0.0002	32	-1
igual	0.0012	0.0006	0.0006	0.0005	0.0008	0.0003	32	-1
mundo	0.0003	0.0006	0.0003	0.0007	0.0005	0.0012	32	1
objetivo	0.0034	0.0009	0.0005	0.0000	0.0000	0.0001	32	-1
realidad	0.0008	0.0013	0.0011	0.0009	0.0002	0.0001	32	-0.625
hizo	0.0002	0.0003	0.0006	0.0016	0.0006	0.0006	31	0.25
memoria	0.0003	0.0005	0.0003	0.0005	0.0009	0.0007	31	0.625
pasar	0.0015	0.0006	0.0006	0.0007	0.0005	0.0002	31	-1
ve	0.0010	0.0006	0.0008	0.0012	0.0003	0.0004	31	0.25
viene	0.0002	0.0001	0.0015	0.0004	0.0007	0.0007	31	-0.25
demás	0.0002	0.0001	0.0003	0.0005	0.0008	0.0011	30	1
experiencia	0.0002	0.0003	0.0012	0.0014	0.0005	0.0003	30	0.25
hice	0.0002	0.0007	0.0008	0.0009	0.0005	0.0006	30	0.25
hubiera	0.0000	0.0001	0.0006	0.0009	0.0001	0.0018	30	1
médico	0.0007	0.0005	0.0003	0.0009	0.0008	0.0004	30	0.375
niños	0.0002	0.0002	0.0018	0.0012	0.0005	0.0002	30	-0.25
primer	0.0000	0.0008	0.0006	0.0007	0.0009	0.0003	30	0.625
quería	0.0007	0.0008	0.0008	0.0004	0.0008	0.0002	30	-0.625
tanto	0.0008	0.0006	0.0005	0.0014	0.0004	0.0004	30	0.25
empezar	0.0007	0.0007	0.0009	0.0007	0.0001	0.0007	29	-0.25
estás	0.0020	0.0005	0.0005	0.0004	0.0005	0.0001	29	-1
mayor	0.0002	0.0000	0.0003	0.0016	0.0009	0.0005	29	0.25
medio	0.0007	0.0006	0.0005	0.0004	0.0008	0.0005	29	0.625
nadie	0.0003	0.0013	0.0005	0.0002	0.0006	0.0004	29	-0.625
pienso	0.0008	0.0016	0.0002	0.0004	0.0002	0.0004	29	-0.625
todavía	0.0007	0.0003	0.0005	0.0007	0.0008	0.0005	29	0.625
época	0.0000	0.0006	0.0009	0.0000	0.0008	0.0007	29	-0.25
aunque	0.0003	0.0009	0.0005	0.0000	0.0005	0.0008	28	-0.625
buena	0.0003	0.0008	0.0003	0.0005	0.0008	0.0004	28	-0.625
dar	0.0000	0.0001	0.0005	0.0005	0.0005	0.0014	28	1
hermano	0.0002	0.0005	0.0003	0.0012	0.0006	0.0006	28	0.25
hermanos	0.0015	0.0007	0.0005	0.0004	0.0005	0.0002	28	-1
pasó	0.0003	0.0001	0.0002	0.0000	0.0004	0.0018	28	1
podía	0.0000	0.0003	0.0000	0.0004	0.0010	0.0009	28	0.625
solo	0.0003	0.0007	0.0006	0.0009	0.0002	0.0007	28	0.25
tuve	0.0002	0.0008	0.0003	0.0009	0.0005	0.0006	28	0.25
chicos	0.0000	0.0003	0.0002	0.0004	0.0005	0.0013	27	1
estas	0.0007	0.0007	0.0005	0.0005	0.0004	0.0006	27	-0.75
hijo	0.0000	0.0001	0.0000	0.0007	0.0012	0.0007	27	0.625
instituto	0.0000	0.0005	0.0025	0.0002	0.0003	0.0002	27	-0.25
marido	0.0002	0.0001	0.0003	0.0011	0.0006	0.0008	27	0.25
mucha	0.0000	0.0006	0.0005	0.0012	0.0006	0.0004	27	0.25
muchísimo	0.0000	0.0002	0.0006	0.0005	0.0010	0.0005	27	0.625
quien	0.0007	0.0006	0.0006	0.0004	0.0004	0.0007	27	NA
seminarios	0.0002	0.0000	0.0017	0.0005	0.0008	0.0002	27	-0.25
tenido	0.0002	0.0005	0.0006	0.0005	0.0008	0.0004	27	0.625
alguien	0.0003	0.0008	0.0005	0.0011	0.0004	0.0003	26	0.25
buscar	0.0008	0.0008	0.0005	0.0004	0.0005	0.0003	26	-0.875
difícil	0.0003	0.0006	0.0005	0.0002	0.0008	0.0005	26	0.625
esas	0.0003	0.0006	0.0002	0.0004	0.0006	0.0007	26	1
hablar	0.0007	0.0010	0.0002	0.0002	0.0003	0.0007	26	-0.625
miedo	0.0003	0.0007	0.0009	0.0005	0.0005	0.0003	26	-0.25
muchos	0.0005	0.0003	0.0005	0.0009	0.0007	0.0003	26	0.25
pasado	0.0003	0.0007	0.0008	0.0004	0.0003	0.0007	26	-0.25
situación	0.0007	0.0002	0.0005	0.0009	0.0009	0.0000	26	0.5
cinco	0.0005	0.0002	0.0005	0.0011	0.0002	0.0007	25	0.25
música	0.0000	0.0001	0.0000	0.0000	0.0004	0.0018	25	1
primera	0.0008	0.0003	0.0000	0.0009	0.0004	0.0007	25	NA
saber	0.0005	0.0003	0.0006	0.0000	0.0008	0.0004	25	0.625
siente	0.0005	0.0005	0.0005	0.0000	0.0009	0.0003	25	0.625
cambio	0.0003	0.0005	0.0003	0.0002	0.0007	0.0006	24	0.625

dicho	0.0010	0.0006	0.0014	0.0004	0.0001	0.0001	24	-0.25
estamos	0.0008	0.0006	0.0014	0.0002	0.0002	0.0002	24	-0.25
hago	0.0003	0.0013	0.0000	0.0000	0.0005	0.0004	24	-0.625
lugar	0.0000	0.0002	0.0002	0.0007	0.0006	0.0008	24	1
mirá	0.0005	0.0001	0.0000	0.0007	0.0012	0.0001	24	0.625
país	0.0002	0.0001	0.0011	0.0009	0.0003	0.0006	24	-0.25
ría	0.0000	0.0000	0.0005	0.0000	0.0010	0.0007	24	0.625
sabe	0.0008	0.0003	0.0002	0.0000	0.0003	0.0010	24	1
sola	0.0010	0.0006	0.0000	0.0002	0.0003	0.0007	24	-1
tuvo	0.0000	0.0002	0.0002	0.0007	0.0010	0.0004	24	0.625
chico	0.0002	0.0002	0.0003	0.0005	0.0004	0.0009	23	1
diciendo	0.0005	0.0005	0.0002	0.0009	0.0005	0.0004	23	0.25
mente	0.0003	0.0006	0.0003	0.0009	0.0000	0.0008	23	0.25
meses	0.0000	0.0002	0.0008	0.0014	0.0004	0.0003	23	0.25
padres	0.0002	0.0007	0.0006	0.0005	0.0002	0.0007	23	-0.625
solamente	0.0000	0.0002	0.0011	0.0004	0.0005	0.0006	23	-0.25
tampoco	0.0003	0.0009	0.0005	0.0002	0.0004	0.0004	23	-0.625
venir	0.0002	0.0001	0.0008	0.0009	0.0007	0.0002	23	0.25
amigo	0.0000	0.0002	0.0003	0.0011	0.0008	0.0001	22	0.25
cuanto	0.0005	0.0002	0.0006	0.0002	0.0006	0.0004	22	NA
hablando	0.0002	0.0002	0.0005	0.0002	0.0004	0.0009	22	1
haciendo	0.0003	0.0001	0.0006	0.0005	0.0005	0.0005	22	-0.25
llegar	0.0000	0.0002	0.0003	0.0009	0.0005	0.0006	22	0.25
noche	0.0002	0.0000	0.0000	0.0007	0.0004	0.0011	22	1
sobre	0.0008	0.0005	0.0003	0.0004	0.0002	0.0006	22	-1
tarde	0.0002	0.0002	0.0003	0.0012	0.0005	0.0003	22	0.25
cual	0.0007	0.0006	0.0003	0.0004	0.0003	0.0004	21	-1
cualquier	0.0005	0.0006	0.0003	0.0002	0.0002	0.0007	21	1
fin	0.0008	0.0006	0.0003	0.0000	0.0002	0.0006	21	-1
hacia	0.0003	0.0003	0.0000	0.0002	0.0008	0.0005	21	0.625
has	0.0007	0.0007	0.0008	0.0000	0.0005	0.0000	21	-0.25
pregunta	0.0007	0.0007	0.0006	0.0002	0.0004	0.0001	21	-0.75
problemas	0.0007	0.0003	0.0005	0.0004	0.0005	0.0002	21	-1
aparte	0.0003	0.0003	0.0000	0.0004	0.0006	0.0005	20	0.625
aspecto	0.0002	0.0001	0.0008	0.0004	0.0004	0.0006	20	-0.25
dentro	0.0000	0.0005	0.0002	0.0007	0.0005	0.0005	20	0.25
estábamos	0.0002	0.0003	0.0000	0.0002	0.0007	0.0006	20	0.625
fuerte	0.0005	0.0001	0.0003	0.0009	0.0004	0.0004	20	0.25
ganas	0.0002	0.0003	0.0000	0.0002	0.0000	0.0014	20	1
historia	0.0003	0.0003	0.0000	0.0000	0.0004	0.0009	20	1
je	0.0000	0.0002	0.0009	0.0004	0.0006	0.0002	20	-0.25
semana	0.0005	0.0002	0.0006	0.0002	0.0002	0.0007	20	1
supervisiones	0.0000	0.0000	0.0023	0.0004	0.0001	0.0002	20	-0.25
tenemos	0.0005	0.0003	0.0002	0.0002	0.0008	0.0002	20	0.625
tenés	0.0012	0.0007	0.0003	0.0004	0.0002	0.0001	20	-1
viaje	0.0005	0.0002	0.0000	0.0000	0.0000	0.0014	20	1
amiga	0.0005	0.0003	0.0000	0.0009	0.0005	0.0002	19	0.25
amigos	0.0003	0.0006	0.0006	0.0000	0.0005	0.0001	19	-0.375
diríamos	0.0002	0.0003	0.0006	0.0007	0.0005	0.0000	19	0.25
entendés	0.0017	0.0010	0.0000	0.0000	0.0000	0.0000	19	-1
falta	0.0003	0.0005	0.0003	0.0004	0.0001	0.0007	19	1
grupo	0.0000	0.0000	0.0011	0.0004	0.0007	0.0001	19	-0.25
lindo	0.0000	0.0000	0.0002	0.0004	0.0008	0.0006	19	0.625
mañana	0.0000	0.0005	0.0002	0.0007	0.0001	0.0008	19	1
momentos	0.0000	0.0005	0.0011	0.0000	0.0005	0.0001	19	-0.25
niño	0.0002	0.0005	0.0011	0.0002	0.0003	0.0002	19	-0.25
podés	0.0005	0.0010	0.0000	0.0002	0.0002	0.0003	19	-0.625
poner	0.0005	0.0005	0.0006	0.0004	0.0003	0.0002	19	-0.25
suerte	0.0012	0.0003	0.0000	0.0000	0.0002	0.0006	19	-1
único	0.0005	0.0007	0.0002	0.0000	0.0004	0.0004	19	-0.625
cambiar	0.0003	0.0006	0.0002	0.0000	0.0006	0.0002	18	0.625
cierto	0.0007	0.0007	0.0003	0.0004	0.0002	0.0002	18	-0.75
fatalismo	0.0022	0.0006	0.0000	0.0000	0.0000	0.0000	18	-1
gustaba	0.0000	0.0003	0.0003	0.0004	0.0005	0.0005	18	0.875
importa	0.0005	0.0005	0.0003	0.0000	0.0006	0.0001	18	0.625
luego	0.0005	0.0007	0.0006	0.0000	0.0000	0.0005	18	-0.625
ningún	0.0000	0.0005	0.0002	0.0000	0.0003	0.0008	18	1
normal	0.0005	0.0009	0.0000	0.0000	0.0004	0.0002	18	-0.625
nuestro	0.0002	0.0002	0.0006	0.0005	0.0002	0.0005	18	-0.25
pacientes	0.0000	0.0001	0.0012	0.0005	0.0001	0.0005	18	-0.25
pareja	0.0008	0.0003	0.0000	0.0004	0.0004	0.0003	18	-1
piensas	0.0003	0.0003	0.0008	0.0005	0.0004	0.0000	18	-0.25

poquito	0.0003	0.0002	0.0008	0.0002	0.0004	0.0003	18	-0.25
primero	0.0003	0.0003	0.0005	0.0007	0.0002	0.0003	18	0.25
supervisores	0.0002	0.0003	0.0012	0.0004	0.0003	0.0000	18	-0.25
tienes	0.0007	0.0003	0.0009	0.0005	0.0001	0.0001	18	-0.25
tuvimos	0.0005	0.0001	0.0002	0.0004	0.0008	0.0001	18	0.625
unos	0.0010	0.0003	0.0002	0.0007	0.0003	0.0000	18	-1
algunas	0.0003	0.0003	0.0002	0.0005	0.0002	0.0005	17	0.25
candidatos	0.0000	0.0002	0.0006	0.0012	0.0003	0.0000	17	0.25
casi	0.0000	0.0007	0.0002	0.0002	0.0005	0.0003	17	-0.625
diga	0.0003	0.0007	0.0002	0.0007	0.0001	0.0003	17	NA
encima	0.0003	0.0006	0.0002	0.0000	0.0003	0.0005	17	-0.625
estudiar	0.0000	0.0007	0.0002	0.0002	0.0004	0.0004	17	-0.625
hija	0.0000	0.0000	0.0000	0.0000	0.0008	0.0007	17	0.625
hubo	0.0002	0.0001	0.0006	0.0002	0.0005	0.0003	17	-0.25
juntos	0.0000	0.0000	0.0000	0.0009	0.0005	0.0005	17	0.25
mala	0.0019	0.0002	0.0002	0.0002	0.0002	0.0000	17	-1
otros	0.0000	0.0002	0.0008	0.0009	0.0002	0.0002	17	0.25
pone	0.0003	0.0007	0.0002	0.0005	0.0000	0.0005	17	-0.625
punto	0.0000	0.0010	0.0000	0.0004	0.0002	0.0003	17	-0.625
quieren	0.0005	0.0001	0.0003	0.0000	0.0007	0.0002	17	0.625
trabajaba	0.0000	0.0002	0.0003	0.0004	0.0008	0.0000	17	0.625
trabajando	0.0000	0.0005	0.0003	0.0004	0.0004	0.0004	17	-0.625
vi	0.0003	0.0003	0.0002	0.0005	0.0002	0.0006	17	NA
vivir	0.0003	0.0006	0.0005	0.0004	0.0002	0.0002	17	-0.625
analistas	0.0003	0.0001	0.0006	0.0012	0.0002	0.0000	16	0.25
ayer	0.0005	0.0000	0.0002	0.0002	0.0005	0.0005	16	-1
calor	0.0000	0.0017	0.0002	0.0000	0.0000	0.0000	16	-0.625
conmigo	0.0003	0.0002	0.0000	0.0007	0.0004	0.0003	16	0.25
contar	0.0002	0.0005	0.0000	0.0005	0.0004	0.0003	16	0.25
daba	0.0002	0.0003	0.0005	0.0002	0.0004	0.0003	16	-0.25
diferencia	0.0000	0.0003	0.0005	0.0002	0.0004	0.0004	16	-0.25
durante	0.0000	0.0002	0.0006	0.0005	0.0003	0.0003	16	-0.25
elegí	0.0000	0.0003	0.0005	0.0004	0.0005	0.0001	16	0.625
esos	0.0000	0.0005	0.0003	0.0002	0.0001	0.0007	16	1
estado	0.0002	0.0007	0.0003	0.0002	0.0002	0.0004	16	-0.625
fuimos	0.0008	0.0000	0.0003	0.0000	0.0002	0.0006	16	-1
ido	0.0000	0.0006	0.0002	0.0005	0.0000	0.0007	16	1
llama	0.0003	0.0003	0.0002	0.0004	0.0002	0.0005	16	1
mira	0.0010	0.0002	0.0006	0.0002	0.0001	0.0002	16	-1
pensando	0.0007	0.0007	0.0002	0.0002	0.0003	0.0000	16	-0.75
personal	0.0000	0.0006	0.0008	0.0004	0.0002	0.0002	16	-0.25
pueda	0.0007	0.0001	0.0002	0.0004	0.0002	0.0006	16	-1
teníamos	0.0008	0.0001	0.0000	0.0004	0.0004	0.0003	16	-1
tomar	0.0005	0.0001	0.0005	0.0005	0.0003	0.0002	16	NA
unas	0.0005	0.0006	0.0008	0.0002	0.0001	0.0001	16	-0.25
ustedes	0.0002	0.0000	0.0003	0.0007	0.0003	0.0005	16	0.25
abuela	0.0007	0.0010	0.0003	0.0000	0.0000	0.0000	15	-0.625
baile	0.0000	0.0000	0.0002	0.0002	0.0002	0.0009	15	1
capaz	0.0007	0.0007	0.0002	0.0004	0.0002	0.0000	15	-0.75
fatalista	0.0015	0.0007	0.0000	0.0000	0.0000	0.0000	15	-1
final	0.0003	0.0002	0.0002	0.0004	0.0002	0.0006	15	1
fondo	0.0014	0.0006	0.0000	0.0000	0.0001	0.0001	15	-1
gustaría	0.0000	0.0003	0.0005	0.0002	0.0002	0.0006	15	1
lo	0.0002	0.0000	0.0006	0.0007	0.0003	0.0002	15	0.25
murió	0.0002	0.0001	0.0000	0.0005	0.0002	0.0007	15	1
mío	0.0003	0.0002	0.0003	0.0000	0.0004	0.0004	15	0.75
nuestra	0.0002	0.0001	0.0003	0.0004	0.0006	0.0001	15	0.625
sentí	0.0002	0.0002	0.0005	0.0004	0.0005	0.0001	15	NA
totalmente	0.0000	0.0005	0.0000	0.0002	0.0005	0.0003	15	0.625
tratando	0.0002	0.0000	0.0005	0.0000	0.0008	0.0000	15	0.625
tío	0.0007	0.0008	0.0002	0.0000	0.0000	0.0003	15	-0.625
vaya	0.0003	0.0001	0.0006	0.0002	0.0002	0.0004	15	-0.25
bailar	0.0000	0.0001	0.0000	0.0000	0.0002	0.0009	14	1
carrera	0.0000	0.0005	0.0000	0.0004	0.0005	0.0002	14	NA
dinero	0.0003	0.0008	0.0005	0.0000	0.0002	0.0000	14	-0.625
ehm	0.0008	0.0010	0.0000	0.0000	0.0000	0.0000	14	-0.625
general	0.0002	0.0000	0.0005	0.0007	0.0002	0.0003	14	0.25
grandes	0.0002	0.0000	0.0012	0.0000	0.0002	0.0003	14	-0.25
iban	0.0000	0.0000	0.0003	0.0004	0.0002	0.0007	14	1
idea	0.0002	0.0003	0.0002	0.0004	0.0002	0.0004	14	NA
internet	0.0003	0.0010	0.0000	0.0002	0.0001	0.0001	14	-0.625
llevo	0.0002	0.0005	0.0000	0.0000	0.0005	0.0002	14	0.625

médicos	0.0000	0.0003	0.0008	0.0011	0.0000	0.0000	14	0.25
ninguna	0.0003	0.0001	0.0003	0.0002	0.0005	0.0001	14	0.625
paciente	0.0000	0.0001	0.0014	0.0002	0.0002	0.0001	14	-0.25
pasando	0.0005	0.0002	0.0002	0.0005	0.0003	0.0001	14	NA
psicólogo	0.0003	0.0001	0.0017	0.0000	0.0000	0.0000	14	-0.25
salir	0.0000	0.0003	0.0000	0.0004	0.0002	0.0006	14	1
sentís	0.0008	0.0001	0.0000	0.0000	0.0006	0.0000	14	-1
tendría	0.0002	0.0000	0.0003	0.0002	0.0005	0.0004	14	0.625
tenían	0.0005	0.0000	0.0002	0.0004	0.0004	0.0003	14	-1
vacaciones	0.0014	0.0006	0.0000	0.0000	0.0000	0.0001	14	-1
éramos	0.0000	0.0002	0.0002	0.0011	0.0002	0.0003	14	0.25
acerca	0.0000	0.0002	0.0008	0.0005	0.0002	0.0001	13	-0.25
afuera	0.0000	0.0000	0.0003	0.0002	0.0005	0.0004	13	0.625
aprender	0.0000	0.0003	0.0000	0.0007	0.0001	0.0005	13	0.25
carácter	0.0003	0.0009	0.0000	0.0000	0.0002	0.0000	13	-0.625
diez	0.0005	0.0000	0.0003	0.0002	0.0005	0.0001	13	1
espectacular	0.0000	0.0000	0.0000	0.0002	0.0004	0.0007	13	1
estaban	0.0005	0.0000	0.0003	0.0007	0.0002	0.0002	13	0.25
mía	0.0002	0.0002	0.0000	0.0000	0.0005	0.0003	13	0.625
opinión	0.0000	0.0008	0.0005	0.0004	0.0001	0.0000	13	-0.625
principio	0.0002	0.0002	0.0005	0.0007	0.0002	0.0001	13	0.25
sabía	0.0003	0.0002	0.0002	0.0000	0.0005	0.0002	13	0.625
sentía	0.0003	0.0002	0.0003	0.0002	0.0001	0.0005	13	1
teléfono	0.0005	0.0003	0.0005	0.0005	0.0001	0.0000	13	NA
temprana	0.0000	0.0000	0.0000	0.0000	0.0005	0.0007	13	1
tenia	0.0000	0.0001	0.0005	0.0005	0.0002	0.0003	13	0.25
textos	0.0000	0.0001	0.0002	0.0000	0.0002	0.0008	13	1
íbamos	0.0003	0.0001	0.0000	0.0002	0.0004	0.0004	13	0.75
adelante	0.0002	0.0003	0.0005	0.0000	0.0002	0.0003	12	-0.25
allá	0.0003	0.0002	0.0005	0.0000	0.0002	0.0002	12	-0.25
cuento	0.0007	0.0003	0.0000	0.0000	0.0001	0.0004	12	-1
dicen	0.0002	0.0006	0.0000	0.0007	0.0001	0.0001	12	0.25
habían	0.0000	0.0000	0.0000	0.0005	0.0003	0.0005	12	0.25
piensa	0.0002	0.0002	0.0006	0.0002	0.0002	0.0002	12	-0.25
pobre	0.0008	0.0002	0.0002	0.0000	0.0002	0.0001	12	-1
podido	0.0005	0.0002	0.0002	0.0004	0.0001	0.0003	12	-1
acuerda	0.0002	0.0001	0.0003	0.0005	0.0002	0.0001	11	0.25
hablo	0.0008	0.0001	0.0002	0.0004	0.0001	0.0001	11	-1
hacia	0.0000	0.0007	0.0000	0.0004	0.0001	0.0002	11	-0.625
pasaba	0.0002	0.0000	0.0000	0.0002	0.0004	0.0004	11	0.75
pudo	0.0002	0.0000	0.0000	0.0002	0.0005	0.0003	11	0.625