



UNIVERSIDAD AUTÓNOMA DEL ESTADO DE MORELOS

FACULTAD DE COMUNICACIÓN HUMANA

**DISEÑO DE UN SISTEMA DE ANÁLISIS DE REDES LÉXICAS PARA LA  
EVALUACIÓN PERSONALIZADA DEL LENGUAJE EN LA DIVERSIDAD**

TESIS

QUE PARA OBTENER EL GRADO DE

MAESTRA EN ATENCIÓN A LA DIVERSIDAD Y EDUCACIÓN INCLUSIVA

PRESENTA

ROSA ELIA RUBÍ BERNAL

DIRECTOR DE TESIS

DR. ALBERTO JORGE FALCÓN ALBARRÁN

CUERNAVACA, MORELOS JUNIO 2018

UNIVERSIDAD AUTÓNOMA DEL ESTADO DE MORELOS

---

FACULTAD DE COMUNICACIÓN HUMANA  
JEFATURA DE PROGRAMAS EDUCATIVOS DE POSGRADO

**DISEÑO DE UN SISTEMA DE ANÁLISIS DE REDES LÉXICAS PARA LA  
EVALUACIÓN PERSONALIZADA DEL LENGUAJE EN LA DIVERSIDAD\***

TESIS  
QUE PARA OBTENER EL GRADO DE  
MAESTRA EN ATENCIÓN A LA DIVERSIDAD Y EDUCACIÓN INCLUSIVA

PRESENTA  
ROSA ELIA RUBÍ BERNAL

DIRECTOR DE TESIS  
DR. ALBERTO JORGE FALCÓN ALBARRÁN

COMITÉ TUTORAL:  
DR. LEONARDO MANRÍQUEZ LÓPEZ  
DR. EDUARDO HERNÁNDEZ PADILLA

\*Este proyecto fue realizado con el apoyo de una beca PNPC otorgada por CONACyT a nombre de Rosa Elia Rubí Bernal

## **Agradecimientos**

A mi director de tesis, el Dr. Alberto Falcón por compartir sus conocimientos, por la orientación, el seguimiento continuo a mi trabajo, pero sobre todo por la motivación y el apoyo recibido a lo largo de estos años.

A los miembros de mi comité revisor: Dr. Leonardo Manríquez, Dr. Eduardo Hernández, Dr. Uliánov Montaña y Dra. Asela Reig, gracias por sus siempre oportunas observaciones las cuales han contribuido de forma significativa tanto a mi trabajo como a mi crecimiento profesional.

A mi familia, en especial a mis padres y hermanos que siempre han estado a mi lado en cada uno de mis sueños y logros. Gracias por el amor, cariño y apoyo que me brindan siempre. Gracias papás, por el esfuerzo que han realizado para apoyarme a cumplir cada una de mis metas, esto es para ustedes, los amo.

A mi esposo Carlos, que siempre me acompaña en mis sueños y aventuras, gracias por el amor y apoyo incondicional en buenos y malos momentos. Eres una parte fundamental en mi crecimiento personal y profesional, gracias por tu paciencia y comprensión en todo momento, pero sobre todo por formar junto conmigo, el mejor equipo.

A mis amigos del Laboratorio de Comunicación Humana y Cognición, por su apoyo, consejos, por el agradable ambiente de trabajo que me brindaron, por estar en esos momentos de diversión, aprendizaje, por llenarme de alegría a cada momento y por hacer del Laboratorio mi segundo hogar.

A todos ellos, muchas gracias.

## Índice

Resumen	7
Capítulo I. Desarrollo del Lenguaje	11
Diferencias individuales	14
Evaluación del lenguaje	18
Capítulo II. Relaciones Lexicas Semánticas	23
Lexicón	25
Aprendizaje de palabras	29
Capítulo III. Redes léxicas	32
Estructura de la red	35
Relaciones lingüísticas para generar redes	38
Teorías de crecimiento de red	43
Capítulo IV. Antecedentes	47
Planteamiento del problema	52
Objetivo general	59
Objetivos específicos	59
Capítulo V. Método	60
1) Desarrollo del software	60
Análisis de corpora	61
Generación de redes en programa de visualización de redes: GEPHI	64
Generación de palabras aprendibles	65
2) Aplicación del sistema	67
Selección del corpus para el análisis de coocurrencia	68
Selección de palabras para la generación de valores de relación sintáctica semántica entre palabras (palabras objetivo) para formar la red global	68
Generación de red global (red modelo)	69
Relaciones de palabras de vocabularios individuales	69
Generación de red individual por niño con sus valores estadísticos propios	70
Generación de red aleatoria igualando cantidad de palabras dentro del vocabulario del niño	71

Comparación vocabularios misma cantidad pero diferente edad	71
Generación de listas palabras aprendibles	71
Resultados	74
Resultados de análisis de vocabularios individuales	76
Resultados de comparación de vocabularios equivalentes	79
Conclusiones	84
Referencias	88
Anexos	93
Desarrollo del software (Descripción técnica)	93

## Resumen

En el presente trabajo se desarrolló un sistema para evaluar la estructura del léxico de los infantes, tomando en consideración las particularidades del vocabulario de los niños en forma individual. Para dicho propósito se diseñó un software que realiza un análisis de la coocurrencia entre las palabras a partir de los corpora de habla espontánea de infantes, lo que en conjunto con un software de acceso abierto, permite analizar los valores estructurales del vocabulario de los niños. A partir de esto, se obtienen valores estadísticos que permiten formar redes léxicas para conocer la organización de las palabras del repertorio léxico de los infantes, lo que permite contar con una evaluación basada en la organización de las palabras de los niños, más que en la cantidad de ellas. Adicionalmente, el sistema creado, tiene el potencial de ser aprovechado para facilitar y fomentar el desarrollo lingüístico a partir de las particularidades en el conocimiento léxico, con una perspectiva de que las diferencias entre los niños son una riqueza que debe tomarse en cuenta para estimular su desarrollo.

**Palabras clave:** redes léxicas, software, desarrollo lingüístico, aprendizaje de palabras

## **Diseño de un sistema de análisis de redes léxicas para la evaluación personalizada del lenguaje en la diversidad**

En diferentes ámbitos, desde el clínico hasta al escolar, se ha visto que el desarrollo del lenguaje es una fuente importante para contar con una referencia sobre el estatus del desarrollo general de una persona. Por tal motivo, la determinación de un retraso en cuanto a la comprensión y producción del lenguaje es fundamental. Los déficits en el lenguaje suelen estar asociados con una debilidad en las habilidades académicas posteriores relacionadas con el lenguaje tales como el desarrollo del vocabulario, la gramática, la memoria verbal y la comprensión de la lectura (Rescorla 2005 cit. en Jiménez 2007).

En ocasiones cuando los padres y educadores observan la variabilidad que hay en el desarrollo del lenguaje entre los niños, optan por realizar una evaluación para cerciorarse que el desarrollo lingüístico es apropiado de acuerdo a la edad del niño o si presenta alguna dificultad. Sin embargo, además de que existen pocas pruebas para conocer o evaluar el desarrollo lingüístico en edades tempranas; la mayoría de las veces estos análisis están basados únicamente en una cuestión cuantitativa refiriéndose con esto al número de palabras adquiridas.

Por otro lado, sabemos que la evaluación del lenguaje, en específico el léxico de una persona no es solamente una cuestión de cantidad, sino

también de la diversidad, del tipo de palabras que se conocen y también de la forma en que se relacionan y organizan esas palabras, lo cual brinda una referencia sobre el almacenamiento y la recuperación de la información. Por ejemplo, podemos pensar en dos niños que tienen vocabularios similares en cantidad, pero podemos hacer una evaluación personalizada o distinta para considerar qué palabras conocen y cómo es que esas palabras se relacionan unas con otras, obteniendo así información diferente y más detallada. Esta información puede ser clave para otros propósitos, como sugieren Styles & Plunkett (2011), que el conocimiento de las relaciones entre palabras es clave para promover el aprendizaje de nuevas unidades léxicas.

En el presente trabajo se desarrolló de un sistema para evaluar la estructura lingüística del léxico de los infantes, basada en las relaciones semánticas entre las palabras, y tomando en consideración las particularidades del vocabulario de los niños de forma individual. Así el repertorio léxico está representado en forma de una red con valores estructurales propios e individuales, independientemente de la cantidad de palabras, lo que permitirá contar con información más completa sobre el desarrollo lingüístico y potencialmente facilitará y promoverá el aprendizaje de nuevas palabras. El sistema de evaluación diseñado, se basa en el principio de que al aprender nuevas palabras, cada niño/niña tiene conocimientos previos distintos, recibieron diferente estimulación en el ambiente de aprendizaje de palabras, cuentan con diversas habilidades,

características y estrategias de aprendizaje que pueden ser aprovechadas a favor de su propio desarrollo y que, por lo tanto es importante diseñar herramientas de evaluación e intervención que tomen en cuenta la amplia diversidad de dichas características y habilidades.

Atendiendo a estas consideraciones, el presente trabajo está conformado por cinco capítulos. El primer capítulo se centra en la adquisición y desarrollo léxico desde los primeros meses de vida del infante, en los indicadores que se toman en cuenta a la hora de evaluar el lenguaje y su importancia en edades tempranas. Dentro del mismo capítulo se detallan también las diferencias entre los infantes con desarrollo típico y los infantes con un vocabulario limitado, considerados hablantes tardíos.

El capítulo II abordará las diversas preguntas acerca de cómo están representadas las relaciones existentes entre palabras y sus implicaciones en el aprendizaje de nuevas palabras.

El capítulo III trata sobre el estudio del lenguaje a través de las redes léxicas, cómo se forman dichas redes y cuáles son sus propiedades estructurales y dinámicas. En el capítulo IV, se presentan diferentes antecedentes relevantes acerca del estudio del lenguaje a través de las redes y las ventajas de analizar y observar el lenguaje desde esta perspectiva, así como los objetivos del presente estudio.

Por último, en el capítulo V se describe el método sobre el diseño y aplicación del sistema.

## **Capítulo I. Desarrollo del Lenguaje**

La adquisición y desarrollo del lenguaje, aunque pareciera ser una tarea sencilla, es en realidad un proceso complejo ya que requiere explicarse con base en diferentes adquisiciones: la conformación fonológica, las reglas morfológicas, la estructura sintáctica y del discurso, el uso pragmático del lenguaje y el léxico (Bloom, 1993). En el presente capítulo nos ocuparemos básicamente de la adquisición del léxico, es decir, la adquisición de palabras.

Mucho se ha documentado de las edades en la que aparecen las primeras palabras, dada la importancia que tiene este logro para el desarrollo posterior. Dichas edades o etapas se presentarán a continuación.

Los infantes típicamente empiezan a asociar secuencias sonoras con significados hacia el final del primer año. Pero anteriormente, a los ocho meses en promedio, muchos niños responden apropiadamente a unas 10 frases familiares, como "¿Dónde está Papi?" (e.g., girando y gateando hacia donde se encuentra papá), o "¡Es hora de bañarse!" (y el bebé se quita los zapatos) (Traxler & Gernsbacher, 2006). Aunque podría ser tentador asumir que el niño está realmente interpretando cada una de las palabras en estas frases, es más probable que los niños estén utilizando una variedad de señales, tanto lingüísticas como contextuales, para dar sentido a estas

expresiones frecuentes (e.g., papá sólo a la izquierda de la habitación, mamá está sosteniendo una toalla junto y juguetes de baño). Sin embargo, el hecho de que los niños respondan de estas maneras indica que están prestando atención al discurso que les rodea y empiezan a entenderlo mediante la asociación de ciertos patrones de sonido en contextos particulares.

Poco tiempo después, los niños empiezan a demostrar una habilidad para entender palabras individuales con un apoyo cada vez menos contextual, una habilidad que seguirá mejorando en los próximos meses. Basándose en los informes de más de 1000 padres, (Fenson et al. 1994 cit en Traxler & Gernsbacher, 2006) se sabe que el nivel medio de 10 meses de edad comprenden aproximadamente 40 palabras, mientras que el nivel medio de 18 meses de edad comprende más de 250 palabras, un aumento de más de seis veces.

En cuanto a la producción de palabras reconocibles, en promedio comienza justo antes del primer cumpleaños del infante. Estas primeras palabras pueden pertenecer a una gran variedad de categorías lingüísticas, pero son generalmente nombres para los cuidadores (e.g., mamá), objetos comunes (e.g., pelota, zapato), expresiones sociales (e.g., bye-bye), entre otras (Nelson, 1973).

Las nuevas palabras tienden a agregarse en los vocabularios expresivos de los niños durante los próximos meses a un ritmo relativamente lento pero constante, alcanzando un promedio de 300 palabras a los 24 meses y más de 60,000 en el momento en que se gradúan de la escuela secundaria (Fenson et al., 1993). Un periodo importante que ocurre durante este crecimiento de aprendizaje de palabras, surge entre los 18 y 20 meses, cuando la velocidad del aprendizaje de palabras aumenta drásticamente, y empiezan a surgir con cierta coherencia las palabras funcionalmente gramaticales, tales como artículos y preposiciones. A este periodo se le conoce como *explosión del vocabulario*, y ocurre una vez que el vocabulario ha alcanzado alrededor de 200 palabras. Algunos investigadores han asociado esta *explosión* con el logro de hitos lingüísticos, por ejemplo, la nueva comprensión que tienen los niños acerca de qué son las palabras (el "naming insight") (Dromi, 1987; Bloom, 1973), habilidades mejoradas de segmentación de palabras (Plunkett, 1992), o habilidades mejoradas para la recuperación de palabras (Dapretto & Bjork, 2000).

Otros investigadores han asociado aumentos en el crecimiento léxico con avances cognitivos relacionados con la naturaleza de la organización de conceptos de objetos (Gopnik & Meltzoff, 1987). Aunque las primeras palabras de los niños provienen de una gran gama de categorías léxicas (Bloom 1973), éstas primeras palabras son típicamente palabras de contenido (sustantivos, verbos, adjetivos), y hasta después aparecen las

palabras de funciones gramaticales tales como preposiciones, determinantes, y pronombres. Dentro de las palabras de contenido, las primeras palabras de los niños tienden a ser referenciales (sustantivos concretos), y solamente más adelante los niños expresarán términos predicativos (e.g., verbos y adjetivos) (Bates, Bretherton, & Snyder, 1988)

Como ya se mencionó anteriormente, existen etapas o hitos de desarrollo en cuanto a la adquisición y desarrollo del vocabulario que corresponde generalmente a la mayoría población infantil, sin embargo, existe una gran variabilidad de adquisición y desarrollo entre niños, sin que estos presenten algún déficit o trastorno. Estas diferencias en el desarrollo se pueden observar desde edades tempranas y puede deberse a diferentes factores, entre ellos la estimulación que reciben en casa, el grado de estudio de los padres así como su nivel socioeconómico, edades de los padres etc. Sin embargo, hay ocasiones en que estas diferencias entre un niño y otro, no se deben a ninguno de estos factores, lo que ha llevado a diferentes investigadores a sugerir que puede deberse habilidades propias del niño para categorizar palabras, entre otras causas. Estos factores se comentarán a continuación.

### **Diferencias individuales**

Anteriormente se reportó el periodo de adquisición y desarrollo del vocabulario dentro del cual se describe la tasa de aprendizaje de nuevas

palabras, que es lo que generalmente corresponde a una parte de la población infantil; sin embargo, existe una gran variabilidad dentro de estos parámetros. Estas variaciones en el tamaño del vocabulario pueden deberse a diversos factores, como el estado y función del aparato anatómico y sistema nervioso, las condiciones de educación, las características del lenguaje que rodean al niño, etcétera. (Navarro, 2003).

En algunas ocasiones las diferencias en el desarrollo léxico pueden observarse, aún entre niños que tienen una historia peri y postnatal sin factores de riesgo, que de igual manera no presentan complicaciones en áreas como la audición, habilidades cognitivas, ni alteraciones neurológicas o en su aparato neuromotor. Cuando existen tales condiciones de desarrollo y aun así se observan limitaciones notorias en la producción del lenguaje, es decir, que no producen 50 palabras específicamente a los 24 meses ni combina frases, es considerado un niño hablante tardío (Jackson-Maldonado, 2004).

El término de hablante tardío ha sido asiduamente utilizado en investigaciones sobre adquisición del lenguaje, y es manejado como sinónimo de "retraso inicial del lenguaje", el cual es caracterizado por presentar al comienzo una atención conjunta más limitada, en su balbuceo producen menos tipos de consonantes y estructuras silábicas, y lo que claramente los diferencia es que empiezan a producir palabras tardíamente.

Se ha encontrado que muchos de los niños con TEL (Trastorno Específico del Lenguaje) (entre el 40% y el 60%) tuvieron un retraso inicial del lenguaje (hablantes tardíos) y la intervención fonoaudiológica podía haber empezado a una edad más temprana (Jackson-Maldonado, 2004). Sin embargo, es importante aclarar que no todos los niños que presentaron retraso inicial del lenguaje presentan posteriormente un trastorno específico del lenguaje, lo que genera más preguntas acerca del origen de estas diferencias y sus implicaciones en el desarrollo posterior.

Los problemas en la comprensión lingüística o los retrasos en la producción adecuada para la edad pueden además derivar otras dificultades posteriores, que pueden ir desde problemas en la comprensión lectora hasta problemas de integración social (Jiménez, 2007). Por ello, es importante que se detecten adecuada y tempranamente, esto sin olvidar que el retraso de lenguaje es un factor de riesgo, más no un trastorno en sí, por lo que existe la necesidad de determinar cuáles son más indicativos de un trastorno posterior.

Aunado a que el retraso simple de lenguaje representa un factor de riesgo, diversos estudios se han enfocado en analizar esta diversidad dentro del desarrollo lingüístico en edades tempranas, debido a que el estudio de los aspectos iniciales de la comunicación de estos niños puede ayudar a entender aún más, cuáles son los componentes iniciales de la comunicación que son fundamentales y predictivos de un desarrollo posterior. Una de las

características principales de un retraso simple del lenguaje, es una menor cantidad de palabras adquiridas a cierta edad. Por ejemplo Colunga y Sims (2011), reportan que a los 2 años de edad, un infante puede llegar a tener adquiridas 300 palabras mientras que otro de exactamente la misma edad solamente 10.

Otros autores como Rescorla, Mirak, y Singh (2000) encontraron que el vocabulario expresivo aumentaba de manera lenta en la población de los considerados hablantes tardíos a comparación de los hablantes regulares; estos últimos pueden tener un crecimiento de 347 palabras entre los 20 y 30 meses de edad, mientras que un hablante tardío adquiere solamente 57 palabras en este mismo tiempo (Jackson-Maldonado, 2004). Este estudio sugiere que las diferencias en el desarrollo pueden indicar que no es sólo un desfase inicial, sino un impedimento persistente que trae consigo débiles habilidades del lenguaje posterior que se reflejan a nivel académico.

Por otro lado, en pruebas para identificar su coeficiente intelectual entre hablantes típicos y hablantes tardíos, se ha encontrado que obtienen puntuaciones típicas e iguales a las de los hablantes regulares (Paul, Looney, & Dahm, 1991). Asimismo se han descartado diferencias relacionadas con el coeficiente intelectual. La variabilidad encontrada en la adquisición del léxico desde edades tempranas, puede entonces estar relacionada con otros aspectos, como la falta de estimulación en el ambiente del niño. Sin embargo, una estimulación pobre del lenguaje no es siempre el factor que

desencadena éste ni otros problemas del lenguaje (Auza, 2009). Esta variabilidad puede también sugerir diferencias en habilidades o estrategias propias del infante para categorizar o aprender palabras, entre otras. Por lo que se debe contar con otros métodos o estrategias para analizar y observar el desarrollo lingüístico en edades tempranas que puedan ser informativas sobre estos diferentes aspectos.

### **Evaluación del lenguaje**

Son diferentes los objetivos por lo que resulta importante evaluar el lenguaje desde edades tempranas. Contar con los instrumentos adecuados de evaluación nos permitirá comprender el desarrollo del mismo, para detectar retrasos o desfases y contar las herramientas necesarias para potencializar las habilidades de cada niño, para a su vez intervenir y estimular de manera oportuna.

Existen aproximadamente 400 pruebas psicométricas que evalúan algún aspecto normativo o de deterioro del lenguaje, ya sea escrito, hablado, fonología, comprensión, lectoescritura, narración, motricidad, bilingüismo, etcétera; (Fuente Buros Institute of Mental Measurement). De los cuales al menos 16 son referidos a la lengua española. Sin embargo, muy pocos abarcan edades tempranas del desarrollo.

Por ejemplo el Peabody Test de vocabulario en imágenes (PPVT-III), tiene como objetivo evaluar el nivel de vocabulario receptivo y el nivel de adquisición de vocabulario de una persona de los 2 años en adelante. Otros

ejemplos de las pruebas utilizadas en el ámbito clínico y de investigación es la Prueba de Lenguaje Oral Navarra (PLON), la cual se aplica a niños de entre 4 y 7 años. Dicha prueba evalúa el lenguaje oral del niño y los aspectos fonológicos, morfológicos, pragmáticos y léxicos en las primeras etapas de la escolaridad (forma, contenido y uso).

Uno de los instrumentos de evaluación del desarrollo lingüístico en bebés más utilizado por psicólogos, terapeutas e investigadores, son los reportes parentales, los cuales consisten en cuestionarios aplicados a los padres o en su caso cuidadores primarios. Tales reportes han representado un método eficaz para identificar tanto lo que comprenden, como lo que producen los infantes desde edades tempranas. Se ha encontrado que estos informes son útiles para la detección de niños con retrasos en el lenguaje, se pueden usar tanto para aplicaciones clínicas como para fines de investigación. Uno de los inventarios más usados a nivel mundial, es el *Inventario MacArthur-Bates del Desarrollo de Habilidades Comunicativas CDI*, el cual se ha modificado y validado en varios idiomas, como es el Inglés y Español de México. Este se divide en dos versiones según la edad del niño, el Inventario I está elaborado para usarse con niños de 8 y 18 meses, mientras que el inventario II es para infantes de 16 a 30 meses (Jackson-Maldonado et al. 2005). Este instrumento es considerado uno de los más confiables para ser utilizados en edades tempranas, debido que se han hecho estudios para corroborar si lo que los padres llegan a contestar en el

inventario del CDI son respuestas certeras y fiables. En una investigación (Karmiloff & Karmiloff-Smith, 2005) pidieron a un grupo de padres responder el CDI y sometieron a los niños a ambientes experimentales con el propósito de corroborar la información que fue proporcionada por los padres. Al realizar las correlaciones entre el *Inventario MacArthur-Bates* y las pruebas experimentales, se encontró que fueron excelentes, lo cual comprueba la confiabilidad que existe en este tipo de reporte parental.

Con todo y que el CDI ha mostrado tener una validez y una confiabilidad significativa, algunos estudios han puesto en cuestionamiento su utilidad en la identificación del riesgo en los déficits y retrasos en el lenguaje (e.g. Feldman, Dollaghan, Campbell, Kurs-Lasky, Janosky & Paradise, 2000). La mayoría de los investigadores reconoce que en el desarrollo del lenguaje observado en los infantes, de por sí, supone una alta variabilidad a lo largo del tiempo. Tal variabilidad dificulta la capacidad predictiva del instrumento por sí solo, sobre todo en trastornos que pudieran estar implicados.

Con base en el estudio recientemente mencionado se puede reconocer la necesidad de hacer aún más robusto el instrumento agregando información que permita reunir más elementos en el análisis del desarrollo lingüístico de los infantes. Información de fondo, como el tipo de palabras en un vocabulario, la relación existente entre ellas y cómo estas impactan o influyen en el crecimiento del vocabulario. Por lo tanto, es importante

generar y mejorar los instrumentos de evaluación actualmente utilizados que brinden información más diversa e informativa del desarrollo lingüístico de los infantes.

Las investigaciones que generan información sobre la estructura del vocabulario son una alternativa para agregar información que pudiera ser clave para explicar el desarrollo lingüístico de los niños. Una de estas investigaciones fue la realizada por Beckage, Smith y Hills (2011), en donde analizaron redes semánticas entre niños hablantes típicos (TD, por sus siglas en inglés *Typically Developing Child*) y hablantes tardíos (LT, por sus siglas en inglés *Late Talker*). En dicho estudio encontraron que las palabras que conforman las redes de los niños TD están conformadas por palabras con relaciones fuertes (cercanas) como: mesa-silla, y las palabras de toda la red cuentan con un mayor grado de relación semántica entre sí. En cambio, las redes de niños LT están formadas por palabras con relaciones lejanas e incluso débiles, por ejemplo: pastel-tijeras. Lo que plantea que los niños LT están aprendiendo y por ende organizando palabras de diferente manera, con un grado menor de relación semántica, que los niños TD. Estos estudios dan pie a observar y analizar cómo se están organizando las palabras en el léxico de los niños, es decir, hacer un análisis de la organización de las palabras del repertorio léxico de los infantes para conocer acerca de la naturaleza de éstos déficits.

Además, como sugieren actualmente algunos autores, las relaciones entre las palabras permiten al hablante acceder a muchas palabras más (Styles & Plunkett, 2011). Este último supuesto se explicará más detallado en los siguientes capítulos, los cuales tratan de las relaciones entre palabras y cómo se generan, además cómo dichas relaciones forman una estructura propia que impacta en el procesamiento de dicha información.

## Capítulo II. Relaciones Léxicas Semánticas

Las relaciones semánticas entre palabras, han captado el interés de filósofos, psicólogos, lingüistas, científicos computacionales, psicólogos cognitivos; en resumen, investigadores de casi cualquier campo cuyos intereses impliquen palabras, significado o la mente. Esto debido a que entre otros temas, existen diversas preguntas acerca de cómo están representadas las relaciones existentes entre palabras. Una de las preguntas más relevantes sobre las relaciones semánticas es, ¿En qué bases se hacen los juicios de la relación semántica?, es decir; ¿Cómo los individuos saben (o determinan) si las palabras están relacionadas semánticamente o no y, si están relacionadas, qué tipo de relación está involucrada?

Existen diferentes perspectivas encargadas de explicar el fenómeno de las relaciones semánticas, ejemplo de ellas es la perspectiva tomada de la pragmática y la perspectiva psicolingüística. Por psicolingüística, se pretende decir que el objetivo es proporcionar un modelo psicológicamente plausible de los conocimientos y procesos involucrados en los fenómenos de relaciones semánticas en el comportamiento del lenguaje humano (Murphy, 2003). Entonces, los modelos estructuralistas, formales y computacionales del léxico se encuentran entre los considerados aquí, los cuales evalúan y modelan las habilidades del lenguaje humano de una manera mental. En cambio, por pragmática, se entiende que los fenómenos lingüísticos son considerados en

referencia a su uso y su estatus en una mente humana dentro de una cultura.

Entonces, como se discutirá a continuación, las relaciones semánticas entre las palabras pueden depender de más que las cualidades semánticas de una palabra, y son altamente dependientes del contexto.

Tiene sentido usar el término relaciones semánticas para indicar relaciones definidas por paradigmas semánticos – pero no antes de definir algunas advertencias/excepciones:

Las relaciones semánticas a veces son usadas para denotar relaciones en frases u oraciones, como implicaciones y contradicciones, pero aquí se debería entender que significa “relaciones paradigmáticas” entre palabras. Dada la perspectiva pragmática tomada aquí y el hecho de que los factores no semánticos pueden afectar a estas relaciones semánticas así llamadas, se podría argumentar que deberían llamarse relaciones pragmáticas. Sin embargo, el término relación léxica se utiliza aquí para indicar cualquier relación paradigmática entre las palabras, no sólo una relación semántica. Así, las relaciones léxicas incluyen relaciones fonéticas (como la rima), relaciones morfológicas (como la variación inflexional), y relaciones morfosintácticas (Murphy, 2003).

Entonces, el término relaciones léxicas puede parecer algo ambiguo, ya que podría referirse a las relaciones entre las palabras (en una página, en

una mente, o dondequiera que ellos pudieran existir) o a las relaciones (entre elementos léxicos) dentro del léxico mental. Para algunos autores, los dos significados son intercambiables, ya que sostienen (o asumen) que si las palabras están relacionadas, entonces esa relación está representada en el léxico (Borovsky, Ellis, Evans, & Elman, 2016).

Una pregunta básica que aún no ha sido contestada es: ¿Dónde se sitúan las relaciones semánticas en un modelo de competencia lingüística y de uso del lenguaje? Muchos creadores de modelos mentales proponen, o asumen, que las relaciones semánticas entre las palabras deben estar representadas en el léxico, con otro conocimiento de las palabras (Murphy, 2003). Por lo que es necesario, tener una idea clara de lo que es el léxico y de lo que contiene, así como de cómo el conocimiento metalingüístico debe y puede ser modelado. La siguiente sección se refiere a esta cuestión.

### **Lexicón**

Si un léxico es una recopilación de información acerca de las palabras, entonces el lexicón mental es la representación mental de la información sobre dichas palabras. Aunque tradicionalmente el lexicón se ha imaginado como un "diccionario de la mente", van surgiendo enfoques más modernos que rompen o reorganizan los límites del lexicón.

Si tomamos, por ejemplo, el modelo representado por Jackendoff (1997; cit. en Murphy 2003), el conocimiento de las palabras se representa

en tres tipos diferentes de conocimiento –Estructuras sintácticas, estructuras fonéticas y estructuras conceptuales– que se intersectan para dar una representación completa del conocimiento que aplicamos con el fin de utilizar la palabra con éxito. Esta estructura divide en tres, previene la relación de palabras en el léxico, puesto que las palabras no se almacenan como entidades distintas.

En resumen, el lexicón es considerado como una estructura compleja organizada en términos de fonología, semántica y sintaxis, entre otros niveles. Cuando se activa una palabra, otras palabras de forma similar, por significado, sintaxis, ortografía, etcétera también se activan, sugiriendo que el lexicón mental es complejo y altamente interconectado (Luce & Pisoni, 1998).

Muchos modelos actuales de procesamiento de lenguaje en adultos proponen que el léxico de adultos funciona como una red de nodos (palabras) vinculados por conexiones a través de las cuales la activación fluye durante el procesamiento lingüístico (Collins y Loftus 1975). Esto es explicado por modelos de activación de propagación que reportan los efectos del priming como una activación que fluye de un estímulo previo (el prime) que complementa la activación ascendente del estímulo de prueba (el target). Cuando dos palabras se presentan en rápida sucesión, el reconocimiento de la palabra objetivo es más rápido (Collins y Loftus 1975),

y más preciso después de una palabra relacionada. El modelo de activación de propagación implica que las relaciones palabra-palabra están así codificadas estructuralmente en la arquitectura de la red, como relaciones entre palabras individuales.

Se propone entonces que el léxico de adultos funciona como un sistema organizado e interconectado explicado por un modelo de activación de propagación, pero surge la pregunta si el léxico infantil es una versión reducida del sistema de memoria semántica en adultos, o si existe una discontinuidad entre las primeras etapas del aprendizaje de palabras y una organización semántica más tardía. En la actualidad, hay poca evidencia acerca de la naturaleza precisa de la organización de léxico infantil, o acerca de cuándo los jóvenes aprendices de palabras empiezan a desarrollar un sistema semántico similar al de un adulto que codifica relaciones tales como asociación, información funcional y categorías.

Uno de los pocos estudios que se han realizado con el objetivo de comprobar si el sistema semántico en edades tempranas es estructurado como el de un adulto, es el llevado a cabo por Hills, Maouene, Maouene, Sheya and Smith (2008), en el cual realizaron el seguimiento del crecimiento de redes de vocabularios, esto a través de los sustantivos del inventario de desarrollo comunicativo de MacArthur-Bates de 20 niños de los 16 hasta los 30 meses. En tal estudio encontraron que uno de los principios de crecimiento de red,

Preferential attachment, sugiere que las palabras que se aprendieron más temprano fueron las que estaban más profundamente integradas en la estructura semántica, prediciendo que los léxicos de edades tempranas serán estructurados semánticamente, y que la organización de la red bien puede ser como la de un sistema adulto, y continuo a través del desarrollo.

Pero, ¿Qué más implica tener un léxico interconectado? Tener un léxico interconectado, al pensar en un gato, por ejemplo, haría más fácil acceder a la palabra *perro*; debido a que los “puentes” entre los significados de las palabras facilitan la transición de información entre las palabras relacionadas (Murphy, 2003).

Además de un mejor acceso a las palabras dentro del vocabulario, una de las ventajas de contar con un léxico organizado e interconectado es que a medida que crece el léxico, y se añade un número creciente de palabras, las representaciones de cada elemento se empiezan a “agrupar” de acuerdo a diferentes criterios, para ser más fáciles de encontrar.

Por ejemplo, al buscar un calcetín en particular en un cajón del guarda ropa organizado de manera aleatoria cuando contamos con un número pequeño de calcetines, resulta una tarea sencilla. Sin embargo, al aumentar la cantidad de calcetines, hacer la búsqueda de uno en específico en un cajón no organizado nos llevaría más tiempo que buscar el mismo calcetín en cajones organizados sistemáticamente, por ejemplo; por colores, o por su

uso, etc.; esto mejoraría y disminuiría el tiempo de búsqueda; es decir, un sistema organizado sería pronosticado para reducir ciertos tipos de procesamiento del lenguaje.

Entonces, si las representaciones semánticas de las primeras palabras aprendidas (edades tempranas) codifican relación semántica entre palabras, el sistema podría ser como el de un adulto, desde sus inicios.

Pero detrás de toda esta organización y sistematización del léxico, ocurren distintos procesos que involucran el aprendizaje de palabras, lo que se abordará a continuación.

### **Aprendizaje de palabras**

A simple vista, aprender una nueva palabra pareciera una tarea sencilla. Sin embargo, detrás del aprendizaje de palabras intervienen diferentes habilidades cognitivas sofisticadas que existen para otros propósitos. Éstos incluyen la capacidad de inferir las intenciones de otros, la capacidad de adquirir conceptos, una apreciación de la estructura sintáctica, y ciertas capacidades generales del aprendizaje y de la memoria (Nematzadeh, Fazly, & Stevenson, 2014) (Bloom 2002). Estos son necesarios y suficientes para el aprendizaje de palabras: los niños los necesitan para aprender los significados de las palabras, y no necesitan nada más.

Diversos autores que observan al aprendizaje de palabras desde el punto de vista de la cognición social, argumentan que el aprendizaje de palabras es el producto de la capacidad de los niños para averiguar lo que otras personas están pensando cuando utilizan palabras. Otros autores interesados en sintáctica han hecho una sugerencia similar por el papel de la sintaxis (Fisher, Hall, Rakowitz, & Gleitman, 1994) ,así como quienes hablan de los requisitos cognitivos previos al aprendizaje de palabras se han preocupado con el marco conceptual y los fundamentos lógicos de dicho proceso (Gasser & Smith, 1998). Sin embargo, otros autores como Bloom (2002) sostienen que una explicación completa de cómo los niños aprenden el significado de las palabras requiere de todas estas capacidades.

Esta discusión entre qué capacidades son las más relevantes para el aprendizaje de palabras ha surgido debido a diferentes preguntas, entre ellas ¿Cómo aprenden los niños que la palabra "gato" no se refiere a todos los animales de cuatro patas, y no sólo a Garfield, sino a todos los miembros de una especie en particular? ¿Cómo aprenden los significados de los verbos como "soñar", adjetivos como "buenos" y palabras para entidades abstractas como "bondad" e "historia"? La adquisición del significado de la palabra es una de las cuestiones fundamentales en el estudio de la mente.

La adquisición de sustantivos incluso simples requiere de capacidades conceptuales, sociales y lingüísticas que interactúen de maneras complejas.

Estas interacciones complejas, se pueden observar directamente con la prodigiosa capacidad de aprendizaje de palabras que los infantes desarrollan entre las edades de uno y dos (Bloom 2002 cit. en Styles y Plunkett 2009).

La prodigiosa capacidad de aprendizaje de palabras, se ha adjudicado a que los niños integran cada palabra que aprenden en un sistema semántico complejo, parecido al de un adulto que codifica la relación entre palabras y facilita el crecimiento del vocabulario (Styles y Plunkett 2009).

Entonces, si el lexicón de los niños está siendo organizado como un sistema complejo e interconectado como el de un adulto, una palabra es representada como un nodo en una red y sus propiedades son representadas como vínculos enlazados con otros nodos (palabras), los cuales tienen diferente peso según su relevancia para el significado del concepto o la similitud fonológica, según sea el caso. Tales relaciones se abordarán en el siguiente capítulo llamado redes léxicas.

### **Capítulo III. Redes léxicas**

Una de las propiedades del lenguaje son las relaciones que existen entre las diferentes palabras que lo componen. Estas relaciones pueden representarse en forma de red, donde además de ofrecer información gráfica sobre cómo están siendo organizadas las palabras en el lexicón mental, brindan medidas o valores estadísticos propias de esas red. Estos valores estadísticos de las redes, proporcionan información relevante sobre su estructura, como el grado de agrupamiento de las palabras y otras medidas de conectividad, lo que se ha relacionado en diferentes estudios con la velocidad de procesamiento, así como el reconocimiento y la recuperación de información (Nematzadeh, Fazly, & Stevenson, 2014). Lo cual ha incrementado el interés de estudiar y analizar el lenguaje a través de la estructura de una red.

En los últimos años hemos sido testigos de una evolución en el estudio de las propiedades estructurales y dinámicas de las redes complejas, entendiéndose como redes complejas aquellas estructuras compuestas por muchas partes que interactúan entre sí, cuya estructura se caracteriza por ser irregular, compleja y evoluciona dinámicamente con el tiempo (Perianes-Rodríguez, Olmeda-Gómez, & de Moya-Anegón, 2008).

El interés por estudiar las propiedades de las redes complejas, se ha visto reflejado en los diversos artículos que se han publicado en revistas de investigación científica internacionales de diferentes disciplinas, que abarcan

biología, física, neurología, sociología, medicina, etc; por mencionar algunos ejemplos. Este interés en las redes complejas radica en que nos hemos dado cuenta de que las redes que abundan en la naturaleza, son parte de nuestra vida diaria y se presentan a diferentes niveles de organización. Por ejemplo, algunas redes biológicas que encontramos en el nivel microscópico son las redes de regulación genética, redes de proteínas, redes neuronales, redes metabólicas. Por otro lado, a un nivel de organización mucho mayor, encontramos redes de comunicación e informáticas como la red internet, redes telefónicas, las redes sociales como amistades, colaboradores científicos, propagación de enfermedades, por mencionar algunas (Aldana, 2006). Estas redes son formadas por un conjunto de elementos, los cuales se denominan vértices o nodos, con conexiones entre ellos, llamadas aristas o bordes. Estas conexiones se pueden determinar de distintas maneras, incluso en un mismo conjunto de nodos podemos definir redes diferentes dependiendo de cómo hayamos definido las conexiones, lo cual depende del fenómeno que nos interese estudiar. Por ejemplo, en una red social los nodos son las personas y las conexiones pueden ser los lazos de amistad que existan entre ellas: dos personas están conectadas si son amigos. En la misma sociedad podemos definir las conexiones de forma distinta, por ejemplo, dos personas están conectadas si son familiares. Claramente, la red definida a través de amistades es diferente a la red definida a través de

relaciones de familia, ya que el hecho de que dos personas sean amigas no significa que sean familiares, y viceversa.

Las redes sociales, biológicas, de comunicación; por nombrar algunas que se han mencionado, son redes que abundan en nuestro alrededor siendo tan diferentes en naturaleza y en tamaño, y que a su vez comparten muchas propiedades estructurales similares entre sí, lo que ha motivado a diferentes investigadores a observar cada vez más fenómenos actuales a través del estudio o teoría de redes, brindando una alternativa para comprender y entender mejor cómo es que funcionan dichos fenómenos analizándose a través de la estructura de una red, mediante de modelos matemáticos formulados para entender y explicar las propiedades estructurales (y en algunos casos también las propiedades dinámicas) de las redes complejas (Aldana, 2006).

En diferentes sistemas complejos que han sido objeto de estudio, podemos encontrar además, redes dirigidas o no dirigidas. Una red dirigida es aquella en la que no todas las conexiones son simétricas, es decir, siempre existen conexiones asimétricas (A está conectado con B, pero B no está conectado con A). Y en el caso de las redes no dirigidas, las conexiones entre los nodos siempre son simétricas (Si A está conectado con B, entonces B está conectado a A) (Aldana, 2006).

Entonces, las redes son parte de nuestra vida diaria y se presentan a diferentes niveles de organización. Pero, ¿qué hacemos con las redes? ¿Cómo se estudian y para qué? El estudio general de las redes complejas puede dividirse en dos campos diferentes y complementarios: Estructura y Dinámica. El primer campo de estudio está interesado en determinar las propiedades estructurales (o topológicas) de la red, es decir, en las propiedades que nos dicen cómo están conectados los nodos unos con otros. Y el campo de estudio de la dinámica, trata de los diferentes procesos o principios en los cuales los nodos interactúan, se sincronizan y se añaden nuevos nodos a la red. En el siguiente apartado nos enfocaremos al estudio de las propiedades estructurales de las redes complejas.

### **Estructura de la red**

A pesar de la gran cantidad y diversidad de propiedades que caracterizan a las redes complejas, hay algunas propiedades que son universales: es decir, son comunes a una gran variedad de sistemas complejos. En general, estas propiedades ofrecen información acerca de la composición de la red, el número de nodos que interactúan y el grado de conectividad entre sus componentes, como también información para conocer si la topología de conexión de la red es regular o aleatoria. Estas propiedades son importantes debido a que son las responsables de su

robustez y eficacia, lo que se relaciona con la velocidad de procesamiento de información dentro de la red (Aldana, 2006).

Además, en los modelos de sistemas que presentan un efecto Mundo Pequeño (conocido en inglés como Small World: seis grados de separación entre un nodo y otro) los componentes muestran una mayor velocidad de propagación, potencia computacional, y sincronización. Por ejemplo, las enfermedades infecciosas se propagan más fácilmente cuando su estructura presenta un efecto Mundo Pequeño que en estructuras regulares (Watts & Strogatz 1998). Además, este efecto conocido como mundo pequeño, se ha encontrado también en redes de lenguaje, es decir en redes generadas por las palabras que conforman un vocabulario. Dicho efecto se observa en mayor proporción en redes conformadas por vocabularios reales en contraste con redes conformadas por palabras al azar, debido a que estas últimas eliminan cualquier patrón de crecimiento ordenado (Beckage, Smith, & Hills, 2011).

Algunas de las propiedades más importantes que determinan la estructura (o topología) de una red son el *Grado de un nodo*, el *Coefficiente de agrupamiento* y la *Distancia geodésica*; las cuales se explican a continuación.

1. Grado de un nodo (*Degree*). En una red no dirigida se define el grado de un nodo como el número total de aristas incidentes en dicho nodo y se

denota por  $k_i$ . En una red dirigida se distingue entre grado entrante (*in-degree*) de un nodo como el número total de aristas que apuntan a un nodo en específico y se representa por  $k_{in} i$ , y grado saliente (*out degree*) de un nodo  $i$  como el número total de nodos a los que apunta el nodo  $i$  y se denota por  $k_{out} i$ .

2. El Coeficiente de agrupamiento (*clustering coefficient*). Es una medida del grado en el que los nodos de una red tienen a agruparse entre ellos. Es decir, es la probabilidad de que dos nodos conectados directamente a un tercer nodo, estén conectados entre sí. Por ejemplo, en una red de amistades, es la probabilidad de que dos de mis amigos sean ellos mismos amigos uno del otro.

3. La distancia geodésica (*geodesic distance*). Es el promedio de las distancias más pequeñas entre dos nodos cualesquiera de la red, es decir el número mínimo de "brincos" que se tienen que dar para llegar de un nodo de la red a otro nodo de la red (Aldana, 2006).

En una red, los nodos además de estar conectados también interactúan, y las interacciones pueden dar lugar a fenómenos dinámicos muy interesantes, sin embargo en el presente apartado nos enfocaremos inicialmente al estudio de las propiedades estructurales de las redes complejas, dejando el tema de las propiedades dinámicas para más adelante.

## **Relaciones lingüísticas para generar redes**

Habiendo explicado las principales propiedades estructurales de las redes complejas, es importante mencionar que en años recientes se ha implementado un nuevo movimiento sustancial en la investigación de la red, que mueve el foco de interés lejos del análisis de pequeños gráficos y las propiedades de vértices individuales o bordes dentro de dichos gráficos, centrándose ahora en estadísticas a gran escala y en las diferentes propiedades que proporciona el análisis de estas redes (Boccaletti, Latora, Moreno, Chavez, & Hwang, 2006).

El análisis de las propiedades estructurales de grandes redes, ha sido conducido en gran parte por la disponibilidad de ordenadores y redes de comunicación que permiten que investigadores recopilen y analicen datos a una escala mucho más grande a lo que antes era posible. Donde los estudios solían estudiar redes de tal vez decenas o en casos extremos cientos de vértices, es bastante común ahora ver redes con millones o hasta mil millones de vértices (Newman, 2003).

Con esta nueva perspectiva ha surgido una evolución en el estudio sobre redes complejas, y el lenguaje ha sido involucrado en este proceso, debido a que lenguaje encaja a la perfección como objeto de estudio bajo la estructura de una red. Esto sugiere que los procesos cognitivos que conducen el aprendizaje del lenguaje eventualmente organizan los

componentes lingüísticos en un llamado léxico mental (ML), el cual es descrito como una estructura de red con elementos léxicos interactuando entre sí (por ejemplo representaciones de la palabra). Es decir, los elementos que componen a una red (en este caso léxica), son los vértices o nodos (palabras), con conexiones entre ellos, llamadas aristas o bordes, los cuales corresponden a las relaciones generadas entre dichas palabras de la red por medio de diferentes relaciones lingüísticas, las cuales serán presentadas a continuación.

Las conexiones de palabras que se generan entre sí son multifacéticas o de naturaleza múltiple, y se debe tener en cuenta los diferentes tipos de relaciones lingüísticas que pueden generar estas interacciones entre elementos léxicos o palabras. Por ejemplo, en el modelo de Collins y Loftus (1975), las palabras son organizadas en una red léxica según su escritura y su sonido y que cada palabra es unida a otras palabras similares en ortografía y/o fonología. Fodor (1983) más tarde propuso que la red léxica también contiene relaciones entre palabras que tienden a coocurrir en el lenguaje cotidiano, como *ratón* y *queso*.

Es decir, las conexiones que se generan entre las palabras que forman las redes léxicas, pueden concebirse a través de diferentes criterios, y cada uno suele ser diferente en los aspectos a los cuales se le otorga mayor relevancia. Algunos lo hacen de acuerdo a las características compartidas

entre palabras, por similitud fonológica, por asociación libre, o por la coocurrencia de las palabras dentro de un corpus lingüístico. Esto quiere decir que, las conexiones entre nodos se pueden determinar por diferentes criterios principalmente: (1) asociación libre, (2) por características compartidas, (3) coocurrencia, y (4) por similitud fonológica.

El primer criterio por el cual se generan relaciones o conexiones entre palabras es por medio de la asociación libre, y funciona en relación con la memoria semántica, la cual nos permite acceder a los significados de los conceptos, a la comprensión de esa información y a disponer de otro conocimiento sin tener necesidad de recuperar las experiencias específicas en las que lo obtuvimos (Vivas & Vivas, 2009).

La técnica comúnmente empleada para la recolección de datos para este criterio, denominada asociación libre, consiste en pedirle a los sujetos que externen deliberadamente -ya sea de manera verbal o escrita- la respuesta que les produce el estar expuestos a ciertas palabras estímulo. Esto permitirá reflejar la relación existente entre dos palabras y dicha relación será lo que determinará los enlaces entre palabras (Barrón, 2013).

Otro de los criterios por los cuales se pueden generar redes léxicas, es por características compartidas. En cuanto a la información recolectada para las normas por características compartidas; se les presenta a los participantes un conjunto de conceptos y se les pregunta por las

características que creen son importantes para cada concepto. Las representaciones obtenidas incluyen una lista de características de un concepto, junto con la frecuencia de producción, es decir el número de participantes que enlistó cada característica para cada concepto de la lista. Esta información básica se puede ampliar con varios análisis estadísticos, complementada con medidas adicionales y se utiliza para probar las predicciones con experimentos humanos y modelado computacional (McRae, Cree, Seidenberg, & McNorgan, 2005).

El tercer criterio para generar relaciones entre palabras y concebir una red léxica, es a través del análisis de coocurrencia de las palabras, en el cual se generan conexiones entre palabras asociadas en nivel léxico y semántico, a través de la reiterada ocurrencia entre dos formas de palabras. Por ejemplo si las palabras "perro" y "hueso" con frecuencia se procesan de forma conjunta, se crea entonces un enlace entre ellas.

Muchas coocurrencias se deben a las relaciones sintácticas entre las palabras, algunas otras son debido a expresiones estereotipadas o colocaciones; las palabras que trabajan en conjunto (por ejemplo la frase "take it easy", en Nueva York) (Cancho & Sole, 2001). Se definen así entonces los enlaces como coocurrencias significativas entre las palabras en la misma frase.

Por último, con respecto al cuarto criterio, en las relaciones generadas por medio de la similitud fonológica, se crea vínculo entre dos nodos si las dos palabras son fonológicamente similares (Vitevitch, 2008). Por ejemplo la palabra foco y foca estarían relacionadas, debido a que solo un fonema las distingue, lo que los hace vecinos fonológicos.

Estos diferentes criterios van de acuerdo a la relación lingüística a la cual se le dé mayor relevancia, buscando un análisis más completo de la estructura y las interacciones del lexicón mental, con el objetivo de conseguir una mayor perspectiva sobre la cognición humana y el desarrollo del lenguaje.

En cuanto al desarrollo del lenguaje y su relación que tiene con el crecimiento de las redes léxicas, autores como Borge-Holthoefer & Arenas (2010) sugieren que el lenguaje (más precisamente el lexicón) es un sistema de crecimiento auto-organizado y que ciertas características o propiedades de la red, permiten la interacción entre palabras y a la vez su crecimiento.

El crecimiento del vocabulario, entendido como el crecimiento mismo de la red, tiene que ver con el estudio de las propiedades dinámicas de las redes.

Considerando que el lenguaje evoluciona, las preguntas serían ¿Qué tipo de crecimiento (en el sentido del aumento del repertorio léxico) conduce a un crecimiento auto-organizado? Y ¿Qué puede estar beneficiando o

limitando este proceso de aprendizaje de palabras? Actualmente han surgido más preguntas en torno a este tema, y una de las más relevantes es ¿Por qué los infantes aprenden algunas palabras más tempranamente que otras?

Algunos expertos apuntan a que el aprendizaje de palabras se ve influenciado por los intereses de los infantes y la disponibilidad cognitiva temprana de algunos conceptos sobre otros, pero ha surgido también respuestas en términos de las propiedades estadísticas del medio ambiente de aprendizaje del lenguaje, apuntando a la frecuencia de palabras en el input y en la importancia que tiene el medio ambiente y la forma en que son presentadas las nuevas palabras para que sean aprendidas por los infantes. Estas hipótesis en cuanto al crecimiento del vocabulario, entendido a su vez como el crecimiento de una red, abarca el campo de estudio de las propiedades dinámicas de redes, lo cual se abordará a continuación de acuerdo a tres principales teorías de crecimiento de red (Hills, Maouene, Riordan, & Smith, 2010a).

### **Teorías de crecimiento de red**

Hills et al (2010) sugieren que hay tres principios o teorías que permiten el crecimiento de una red del lenguaje (red dirigida), cada uno de ellos con una dinámica de crecimiento distinta, basándose en los nodos o palabras ya adquiridas y la influencia de las nuevas palabras presentadas en diferentes contextos. Estos principios pueden predecir qué palabras (nodos)

tienen más probabilidad de ser aprendidas antes que otras, y pueden brindar herramientas para promover el aprendizaje de nuevas palabras tomando en cuenta la composición de la red del lenguaje. Cada principio y sus características se explicarán a continuación.

Los principios de crecimiento de red para el aprendizaje temprano de palabras son tres: 1. *preferential attachment*, 2. *preferential acquisition* y 3. *lure of the associates*. (ver fig. 1)

1. El principio de preferential attachment (palabra nueva relacionada a palabra conocida con mayor indegree) está basado en la conectividad de las palabras conocidas con otras palabras conocidas, es decir, sugiere que las palabras que primero son aprendidas son las que se relacionan con palabras ya aprendidas con un alto valor de grado entrante (indegree) obtenido de las mismas palabras conocidas. Por ejemplo, si el infante tiene adquiridas en su vocabulario las palabras: perro, pelota, hueso y croqueta; y pelota, hueso y croqueta, se relacionan es decir apuntan hacia la palabra perro, la palabra próxima a aprender deberá ser una relacionada directamente con la palabra perro, ya que tiene 3 palabras que están apuntando hacia a ella y eso le da un alto valor de grado entrante, lo que sugiere tiende a “atraer” nuevas palabras del medio ambiente de aprendizaje.

2. El principio de preferential acquisition (palabra nueva con mayor indegree) se basa en la conectividad de las palabras en el ambiente

aprendizaje, sugiriendo que las palabras aprendidas más tempranamente son las más diversas contextualmente en el lenguaje aprendizaje. Es decir, que aparecen constantemente en diferentes contextos en el input, por ejemplo si la palabra "plato", aparece constantemente acompañada de diferentes palabras, tendrá ventaja ser aprendida sobre las demás que aparecen en conjunto siempre con las mismas palabras.

3. El principio de lure of the associates (palabra nueva con mayor indegree de palabras conocidas) sugiere que se aprenden nuevas palabras de acuerdo a cuantas palabras conocidas se relacionan con las nuevas palabras por aprender, esto significa que las palabras aprendidas más tempranamente tienen más conexiones con palabras conocidas. Por ejemplo, si se tienen adquiridas las palabras, pelota, hueso, croqueta y perro, las palabras con mayor probabilidad de ser aprendidas eventualmente, será la que tenga más relaciones con dichas palabras del vocabulario, por ejemplo la palabra relacionada: cachorro.

Con base en estos tres principios o teorías de crecimiento de red, se han realizado diferentes estudios tratando de encontrar cuál de ellos encaja más al crecimiento de una red del lenguaje, lo cual se presentará en el siguiente capítulo.

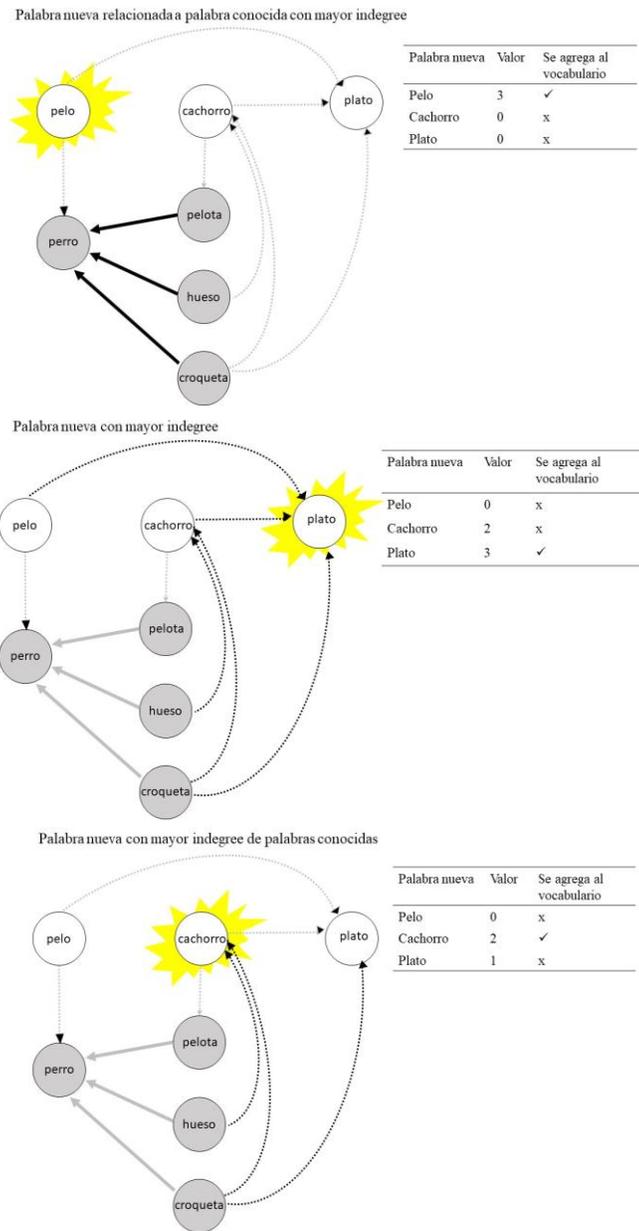


Figura 1. Teorías de crecimiento de red, esquema basado en Hills, Maoene, Riordan, & Smith, (2010)

## Capítulo IV. Antecedentes

Se han realizado diversos estudios para analizar y observar las propiedades estructurales de las redes del lenguaje, como también cuál es el principio de crecimiento de red que corresponde más al crecimiento de estas redes.

En cuanto al análisis de la estructura del lenguaje a través de redes, uno de los estudios que ha analizado la estructura de la red semántica en el léxico de infantes, es el realizado por Beckage, Smith y Hills (2011) en el cual analizan la organización de las palabras en el léxico mental de infantes de 15 y 36 meses, considerando relevante el análisis de la estructura de la red para conocer cómo se está llevando a cabo este proceso y qué relación e impacto tiene sobre el desarrollo lingüístico. En dicho estudio, antes de la edad de adquisición de determinadas palabras y su papel en las redes semánticas, utilizaron datos normativos sobre las palabras que los niños suelen saber a una edad determinada y crearon redes de palabras que se conectan en vocabularios normativos según su relación semántica, de acuerdo a las estadísticas de coocurrencia de las palabras dentro de corpus recogidos de lenguaje escrito o hablado.

Beckage, Smith y Hills (2011) analizan por primera vez la estructura de redes semánticas individuales de niños que tienen un típico o muy pequeño vocabulario para su edad y sexo, y la relación entre las diferencias

individuales en la estructura, como también las diferencias individuales en el ritmo de desarrollo léxico.

Los análisis se enfocaron en los vocabularios de niños hablantes típicos y hablantes tardíos. Un infante es considerado hablante tardío cuando se observan limitaciones notorias en la producción de su lenguaje, es decir, que no producen 50 palabras específicamente a los 24 meses ni combina frases (Jackson-Maldonado, 2004). Además, es importante recalcar que estas diferencias en el desarrollo lingüístico que presentan los niños hablantes tardíos se observan aún y cuando demuestran una historia peri y postnatal sin factores de riesgo, y de igual manera no presentan complicaciones en áreas como la audición, habilidades cognoscitivas, ni alteraciones neurológicas o en su aparato neuromotor. Esta es una de las razones por las que se trata de explicar qué es lo que puede estar determinando el lento o rápido aprendizaje de palabras de un niño a otro.

En el estudio realizado por Beckage et. al (2011), posterior al análisis de las redes semánticas entre niños hablantes típicos y hablantes tardíos, encontraron que las dos redes son muy diferentes en su estructura: la red del niño TD (hablante típico, TD por sus siglas en inglés Typically Developing Child) muestra una mayor conectividad (in degree), una estructura más local (coeficiente de agrupamiento) y el acceso más global (distancia geodésica) que la red del niño LT (hablante tardío, LT por sus siglas en inglés Late

Talker). Esto quiere decir, que una de las grandes diferencias entre redes de niños TD y LT, son las relaciones semánticas que existen entre las palabras que tienen adquiridas, lo que hace que en las redes de niños TD exista una mayor relación semántica entre las palabras que forman la red, ocasionando una rápida accesibilidad a todas las palabras de la red. Por ejemplo las redes de niños TD, están conformadas por palabras con relaciones fuertes (cercanas) como: mesa-silla, y las palabras de toda la red cuentan con un mayor grado de relación semántica entre sí; en cambio, las redes de niños LT están formadas por palabras con relaciones lejanas e incluso débiles, por ejemplo: pastel-tijeras. Lo que plantea que los niños LT están aprendiendo y por ende organizando palabras de diferente manera, con un grado menor de relación semántica, que los niños TD.

Esto puede sugerir que algunos clusters de palabra o incluso ciertas palabras dentro de la red, pueden estar desempeñando un papel fundamental de conexión entre palabras, lo que facilita que se agreguen más o menos palabras a la red, es decir, que ciertas palabras sean clave para que se añadan más palabras a la red. Esto se relaciona con una propiedad característica que pueden presentar algunas redes, las redes del lenguaje por ejemplo, y es un fenómeno conocido como efecto de redes de "*mundo pequeño*" (small world en inglés), dicho efecto permite a altas cantidades de estructura local combinarse con el acceso global de la red. Es decir una red que presenta un efecto mundo pequeño, presenta un agrupamiento muy

elevado entre algunos de sus nodos y con cortas distancias entre pares de nodos. Esta presencia de algunos nodos con muchas conexiones son los llamados "hubs" y funcionan como "súper conectores" los cuales crean atajos que permiten que los recursos y la información salten de grupo a grupo, eliminando la necesidad de seguir caminos largos. Un ejemplo de ello es la teoría de que todos estamos conectados a través de un máximo de seis contactos intermedios, el cual es el origen del término "mundo pequeño" que representa dicho fenómeno. Pero esto hace a la red potencialmente frágil, pues la eliminación de unos pocos de esos "súper conectores" desbarataría la red, que es lo que puede estar pasando en las redes de niños hablantes tardíos en el estudio de Beckage et. al (2011), al estar ausentes palabras claves que funcionarán como "superconectores" o "hubs" la red se ve limitada en su crecimiento, a diferencia de una red de un hablante típico. Además, en el estudio realizado por Beckage et. al (2011), confirmaron que el efecto conocido como mundo pequeño, se observa en mayor proporción en redes conformadas por vocabularios reales en contraste con redes conformadas por palabras al azar, debido a que eliminan cualquier patrón de crecimiento ordenado (Beckage et al., 2011).

En cuanto a patrones de crecimiento se refiere, a su vez otros estudios han indagado en cuál principio o teoría de crecimiento de red que más encaja a la estructura de redes del lenguaje. Por ejemplo, en un estudio realizado por Beckage, Aguilar y Colunga (2015), representaron el léxico como una red

fonológica o semántica y construyeron vocabularios normativos de infantes entre 16 y 30 meses de acuerdo al Inventario MacArthur-Bates del Desarrollo de Habilidades Comunicativas *CDI*. Se utilizó un algoritmo de crecimiento de adquisición preferencial para predecir y cuantificar las palabras que se aprenderán posteriormente y comparar los dos tipos de redes (fonológica y semántica). Este estudio confirmó que la información fonológica puede desempeñar un papel en el aprendizaje de nuevas palabras en edades tempranas o podría sugerir en cambio importante en el desarrollo lingüístico, incluso más acentuado que la información semántica.

En otro estudio realizado por Hills et al., (2010), compararon los tres principios de crecimiento de red para ver cuál encaja mejor con el crecimiento de una red de lenguaje. Los resultados de este estudio indican que las palabras que son el target (palabra blanco) de una larga serie de cues (palabra respuesta); es decir que tiene más asociaciones libres son más propensos a ser aprendidas antes de las palabras asociadas a un conjunto más reducido de "cues". Los resultados corresponden al principio de crecimiento de red *Lure of the Associates*, el cual sugiere que las palabras aprendidas tempranamente que coocurren con otras más palabras en el ambiente de aprendizaje del lenguaje, por consecuencia serán más densas en cuanto a relaciones semánticas formando *hubs* en la red semántica, y como resultado estas palabras que aparecen en diferentes contextos son las más probables en ser aprendidas en cualquier momento determinado.

## **Planteamiento del problema**

El lenguaje puede ser observado y evaluado por medio de diferentes herramientas, sin embargo la mayoría de ellas están diseñadas para edades de 3 años en adelante. Una de las pocas herramientas que brinda la posibilidad de evaluar el desarrollo lingüístico desde edades tempranas son los reportes parentales, los cuales suelen ser cuestionarios que brindan información acerca de cuáles y cuántas palabras tiene adquiridas un infante a una edad determinada. Los resultados obtenidos en dichas evaluaciones son útiles para diferenciar a niños con un desarrollo típico del lenguaje de aquellos que pueden presentar un retraso de lenguaje, sin embargo, no ofrecen información acerca de la naturaleza de dichos déficits del desarrollo del lenguaje, así como información acerca de cómo se pueden prevenir o estimular el desarrollo lingüístico.

Uno de los acercamientos al estudio del lenguaje en edades tempranas que puede brindar información acerca de los mecanismos y procesos de aprendizaje de palabras, así como la capacidad de explicar la naturaleza de diferentes déficits, es analizando el lenguaje a través de la estructura de una red con sus propiedades estructurales.

Existen diferentes estudios que han examinado el papel de la estructura de la red en el léxico mental de niños pequeños durante la adquisición del lenguaje, que además de proporcionar información para comprender el complejo sistema que es el lenguaje, pueden generar

alternativas de estimulación del desarrollo lingüístico desde edades tempranas, propiciando y facilitando el acceso a la mayor cantidad de palabras y potencializar su aprendizaje.

Además, observar el lenguaje a través de la teoría de redes y su crecimiento, permite valorar otras características relevantes del desarrollo lingüístico, no sólo se centrado en la cantidad de palabras que tienen adquiridas los infantes (que es lo que obtiene comúnmente de reportes parentales por ejemplo), sino que trata de ver más allá; analizando las relaciones que existen entre palabras, el tipo de palabras que se conocen, incluso el papel que desempeñan tanto las palabras aprendidas como las características del ambiente de aprendizaje del lenguaje y cómo es que impactan o influyen en el desarrollo del mismo.

Hacer este tipo de análisis del desarrollo lingüístico plantea la importancia de considerar no solo las deficiencias o debilidades del niño, sino, analizar otros factores que toman en cuenta las fortalezas o habilidades, en conjunto con el contexto del ambiente de aprendizaje. Realizar la observación del lenguaje a través de la teoría de redes, a diferencia de los reportes parentales, por decir un ejemplo, ya que presentan diversas ventajas en su evaluación, al ser confiables y al brindar información relevante, están limitados a reportar solamente cantidades de palabras adquiridas, dando poca información sobre las características cognitivas de

los niños. Los resultados arrojados se centran en los déficits, es decir si tienen menos o más palabras adquiridas; dejando dudas sobre lo que está determinando estas diferencias en el desarrollo lingüístico entre unos niños y otros y de cómo se pueden prevenir o intervenir.

Una de las alternativas para conocer más sobre estas discrepancias sería indagar sobre la organización de las palabras en el léxico mental de los niños, lo que parece relevante para conocer cómo se está llevando a cabo este proceso y qué relación e impacto tiene sobre el desarrollo lingüístico. Como el estudio realizado por (Beckage et al., 2011) en el cual posterior al análisis de las redes semánticas de dos grupos de niños hablantes típicos y tardíos, encontraron que una de las diferencias entre grupos son las relaciones semánticas que existen entre las palabras que tienen adquiridas, lo que hace que en las redes de niños hablantes típicos exista una mayor relación semántica entre las palabras que forman la red, ocasionando una rápida accesibilidad a todas las palabras de la red; lo que plantea que los niños hablantes tardíos están aprendiendo y por ende organizando palabras de diferente manera, con un grado menor de relación semántica, que los niños hablantes típicos. Tales resultados, pueden sugerir que algunos clusters de palabra o incluso ciertas palabras dentro de la red, pueden estar desempeñando un papel fundamental de conexión entre palabras, lo que facilita que se agreguen más o menos palabras a la red, es decir, que ciertas palabras sean clave para que se añadan más palabras a la red, lo que ofrece

información relevante que en otros métodos de evaluación no se está tomando en cuenta y que puede ser fundamental para explicar los procesos de aprendizaje de palabras en edades tempranas así como para promover el aprendizaje de nuevas unidades léxicas.

Por lo tanto, es importante partir de la idea que no siempre la cantidad de palabras es el mejor indicador del desarrollo lingüístico de los infantes, sino que es plausible que la organización de dichas palabras del repertorio léxico de los infantes, sea lo que puede estar determinando el lento o rápido aprendizaje de palabras posteriores a las que ya se conocen. Realizar un análisis del desarrollo lingüístico mediante las redes léxicas, puede brindar entonces un panorama más amplio acerca de las características individuales de los niños, centrándose no en los déficits o debilidades, sino, identificando y utilizando el conocimiento y aprendizaje actual para potencializar sus habilidades de acuerdo a las particularidades de su desarrollo. Lo cual permitirá ofrecer un contexto enriquecedor que favorezca su desarrollo desde edades tempranas y así, prevenir futuras dificultades en el desarrollo.

Bajo esta perspectiva del análisis de la estructura lingüística mediante las redes léxicas, se desarrolló el diseño del programa LEXNET-UAEM 1.0.6, el cual consta de un sistema computacional que permite una evaluación de la organización de las palabras del léxico mental de los niños. Como parte de este gran objetivo, el software LEXNET-UAEM 1.0.6 permite el análisis de las

palabras en un corpus de habla espontánea, con medidas cuantitativas sobre las relaciones semánticas de palabras, lo que adite a su vez generar un modelo de red léxica que sirve como parámetro para la evaluación de la organización de las palabras de los infantes.

Para la red de cada niño en particular, la información obtenida a través del software LEXNET-UAEM 1.0.6, permite generar una red por medio de un programa de visualización de redes llamado Gephi. El cual computa tres estadísticas de la red sobre todas las palabras en la red. Estas serán grado entrante, coeficiente de agrupamiento, y la distancia geodésica (*in-degree, clustering coefficient, and geodesic distance*).

En conjunto, estas tres estadísticas proporcionan tres diferentes niveles de información sobre la estructura de la red, lo que permite conocer las particularidades del repertorio del léxico de cada niño, lo cual a su vez tendrá el potencial para crear programas de intervención que tomen en cuenta las diferencias individuales de cada uno de los evaluados, para que atienda el potencial del desarrollo lingüístico de cada niño. De manera tangencial, estos mismos resultados podrán contestar preguntas relacionadas con los factores influyentes en la organización léxica de los niños. Ya que brinda la posibilidad de analizar otros factores y relacionarlos con los resultados obtenidos por el software.

En estudios previos que analizan las diferencias en la adquisición del lenguaje examinando redes semánticas de niños hablantes típicos y hablantes tardíos como el de Beckage et al., (2011), se plantean preguntas acerca de los factores influyentes en la organización léxica de los niños, más específicamente del por qué los niños que tiene un ritmo lento en el aprendizaje de palabras no adquieren conjuntos de palabras que presentan estructuras de red de mundo pequeño. Los autores se han cuestionado si es plausible que el medio de ambiente aprendizaje se esté distorsionado por diferente factores, lo que orilla al infante a recurrir a estrategias diferentes a las que emplean niños que muestran desarrollo típico. En resumen, esta es una posibilidad que debe ser probada mediante el análisis de red de la estructura del habla dirigida a infantes de manera individual, centrando la atención en el contexto del ambiente aprendizaje que rodea al menor.

Otra de las funciones del software LEXNET-UAEM 1.0.6, permite generar predicciones del vocabulario de cada infante con mayor probabilidad de aprendizaje de acuerdo a las teoría del crecimiento de la red; el programa proporciona información acerca de cuáles palabras que no están en el vocabulario de cada niño tienen mayor probabilidad de ser aprendidas posteriormente, de acuerdo a los tres diferentes principios de crecimiento de red para el aprendizaje temprano de palabras (*preferential attachment*, *preferential acquisition* y *lure of the associates*). Esto ofrece información relevante y necesaria para llevar a cabo una estimulación lingüística

oportuna dirigida al aprendizaje de nuevas palabras en función de su importancia.

En conclusión, el diseño y desarrollo del programa LEXNET-UAEM 1.0.6, parte de una visión basada en las fortalezas, no en las debilidades de cada niño, permitiendo aprovechar las diferencias para promover y potencializar el desarrollo lingüístico. Recordando que el concepto de diversidad no se refiere solamente a la capacidad para aprender, y que suele ser medido exclusivamente a través del rendimiento académico por mencionar un ejemplo, sino a los distintos modos y ritmos de aprendizaje de cada niño o niña de acuerdo a su particular modo de ser. Una comprensión de la heterogeneidad supone aceptar la diversidad como elemento de progreso y de riqueza de la colectividad, y por otro lado, asumir las diferencias en las personas como elementos que van a enriquecer su aprendizaje y el de los demás (Melero, 2004).

El diseño del programa LEXNET-UAEM 1.0.6, está basado en el principio de que cada niño/a tiene características, intereses, capacidades y necesidades de aprendizaje distintos y deben ser los sistemas educativos, educadores, padres de familia y herramientas de evaluación e intervención los que deberían de estar diseñados teniendo en cuenta la amplia diversidad de dichas características y necesidades.

## **Objetivo general**

- Crear un sistema para la evaluación de la estructura lingüística de los infantes por medio del análisis de las redes léxicas, que permita observar y analizar las propiedades estructurales de dicha red permitiendo contar con información acerca de la composición de la misma, el número de nodos que interactúan, el grado de conectividad entre sus componentes, y la velocidad de procesamiento de información dentro de la red.

## **Objetivos específicos**

-Diseñar un software que por medio de análisis de coocurrencia de las palabras dentro de un corpus lingüístico seleccionado, arroje valores estadísticos que permitirán formar una red léxica.

-Comparar los valores estadísticos obtenidos de redes de vocabularios reales y redes configuradas con palabras aleatorias.

- Analizar el vocabulario de los infantes obteniendo valores estadísticos sobre su fuerza de asociación, cohesión y cantidad de relaciones entre las palabras que conforman tal léxico de niños de entre 16 y 32 meses

-Generar predicciones del vocabulario con mayor probabilidad de aprendizaje de acuerdo a las teorías de crecimiento de la red.

## **Capítulo V. Método**

El presente trabajo se llevó a cabo de manera interdisciplinaria, colaborando con profesionales en psicología, terapia de lenguaje y profesionales de la programación computacional, esto para lograr un producto que pueda ser posteriormente utilizado tanto como un instrumento de evaluación e intervención, así como de investigación. Dicho proyecto se desarrolló en dos fases: 1) Desarrollo del software LEXNET-UAEM 1.0.6, y 2) Aplicación del sistema. La descripción técnica del desarrollo del software expondrá en el apartado de "Anexos". La descripción del método se realizará en relación a los aspectos teóricos y a los antecedentes del presente trabajo.

### **1) Desarrollo del software**

El software LEXNET-UAEM 1.0.6 diseñado en el presente proyecto, es una aplicación que tiene dos principales objetivos, uno de ellos es el análisis de corpus lingüísticos para generar relaciones sintáctico-semánticas entre palabras, y el segundo es generar valores derivados de diferentes tipos de relaciones léxicas para generar redes léxicas. Esto con el propósito de crear un sistema de evaluación del desarrollo lingüístico aplicable desde edades tempranas, el cual permite obtener información relevante del vocabulario de los infantes, no sólo centrándose en la cantidad de palabras que tienen adquiridas los infantes (que es lo que obtiene comúnmente de reportes

parentales, por ejemplo), sino agregando otros valores estadísticos, que ya se han mencionado anteriormente.

En la evaluación de vocabularios puntuales, el sistema permite analizar las relaciones que existen entre palabras, el tipo de palabras que se conocen y los valores estadísticos de la red individual que estas palabras conforman. Además, en la investigación, el sistema permite explorar el papel que desempeñan tanto las palabras aprendidas como las características del ambiente de aprendizaje del lenguaje y abre la posibilidad de analizar cómo es que impactan o influyen en el desarrollo del mismo.

Para dichos propósitos se desarrollaron diferentes funciones y criterios, las cuales se presentarán a continuación.

### **Análisis de corpora**

Para generar redes léxicas, existen diferentes procesos que deben realizarse con anterioridad, con el objetivo de contar con la información necesaria para formar las relaciones entre las diferentes palabras que conforman la red.

El análisis de coocurrencia en el software permite el análisis de diferentes tipos de palabras, como sustantivos, adjetivos, verbos, etc. Dichas palabras estarán definidas en dos listas ingresadas por el investigador, lo que servirá para identificar las relaciones entre una palabra y otra. A partir de

este análisis se generan conexiones entre palabras asociadas en nivel léxico y sintáctico-semántico, a través de la reiterada ocurrencia entre dos formas de palabras, es decir contabilizando cuántas veces dos palabras coocurren dentro de un texto en un rango de palabras determinado. Por ejemplo si las palabras "perro" y "hueso" con frecuencia se procesan de forma conjunta, se crea entonces un enlace entre ellas.

Además el análisis a partir del software LEXNET-UAEM 1.0.6, se puede ejecutar dentro de uno o varios corpora. Estos corpora pueden ser cualquier archivo de texto, del cual se quieran obtener las relaciones sintácticas-semánticas entre las palabras seleccionadas.

Al utilizar las estadísticas de coocurrencia como el índice de relación semántica, en las redes, la palabra A (Palabra SOURCE) se conectará a la palabra B (Palabra TARGET) si aparece dentro de las primeras cinco palabras (o el tamaño de ventana que se elija) que siguieron a la palabra B, produciendo una red dirigida en el que una relación de A B no implica B también está conectado a A. Este análisis permite contar con una base de datos que contenga las relaciones existentes entre pares de palabras, lo que facilita la creación de una red léxica con sus valores estadísticos propios.

En cuanto al tamaño de la ventana se refiere, es el rango de palabras dentro de las cuales se contabilizará la coocurrencia de una palabra con otra. El tamaño de la ventana podrá elegirse en el sistema, sin embargo el tamaño

promedio utilizado en diferentes investigaciones es de cinco, y se basó en el trabajo previo que investigó el poder predictivo a través de múltiples tamaños de ventana (Hills et al., 2010b).

Posterior al análisis de coocurrencia, el sistema genera una base de datos con las relaciones obtenidas entre todas palabras unas con otras (ver fig. 2).

Una vez que se cuenta con la base de datos con las relaciones que incluyen todas las palabras, se ingresa el vocabulario individual de cada infante, y se obtiene una base de datos con las palabras conocidas y sus valores de relaciones correspondientes, listas para ser ingresadas en el Programa de Visualización de Redes Gephi, para obtener la red gráficamente con sus valores estructurales propios de cada infante.

No.	I.S.	SOURCE	I.T.	TARGET	1o	2o	3o	4o	5o	FR-	SM-	FR+	SM+	FA	INDEGREE	OUTDEGREE	DEGREE
64	19	Conejo											81		22	20	42
65			4	Animal					1	1	0		2	2.47			
66			8	Bicho						1	0		1	1.23			
67			24	Gato	7	3	5	3	2	0			20	24.69			
68			29	Jirafa					1	0			1	1.23			
69			39	Pato					1	0			1	1.23			
70			41	Perro	3	3			1	0			7	8.64			
71			49	Platón			1						1	1.23			
72			51	Tigre	1								1	1.23			
73			117	Manzana			1						0	1.23			
74			135	Pollo	3	6	2	1	2	0			14	17.28			
75			139	Sal	1	1	2		1	0			5	6.17			
76			155	Zanahoria	1	2		1	1	0			5	6.17			
77			239	Muñeca						1	0		1	1.23			
78			388	Casa	2	3		3		0			8	9.88			
79			413	Amiga			1		1	0			2	2.47			
80			415	Bebé			1			1	0		2	2.47			
81			425	Maestra				1	1	0			2	2.47			
82			426	Mamá	1	1			1	0			3	3.7			
83			442	Señora					1				1	1.23			
84			443	Señorita	1	1				1	0		3	3.7			

Figura 2. Reporte global, con los valores de relación entre palabras obtenidos por el análisis de coocurrencia

## **Generación de redes en programa de visualización de redes:**

### **GEPHI**

El programa de visualización de Redes GEPHI, es un software de libre acceso (open-source) de análisis de redes y visualización escrito en Java en la plataforma NetBeans, el cual permite generar redes con sus propiedades estructurales, que marcan las relaciones entre actores o nodos que componen dicha red.

El análisis y visualización de la red a través de GEPHI, es una herramienta útil para visualizar los datos (en este caso vocabulario de infantes) de forma gráfica, lo que permite analizar la información desde un ángulo distinto a los métodos de evaluación tradicionales.

El programa GEPHI, permite generar entonces una red gráfica con características estructurales propias como lo son el grado de un nodo, coeficiente de agrupamiento, diámetro de la red, distancia geodésica, entre otras. Estos valores estadísticos de las redes, brindan información relevante sobre su estructura. Para la obtención de estos valores estructurales, GEPHI trabaja con bases de datos que contiene información del grado de relación entre los actores, en este caso, palabras. Dichas bases de datos son generadas a través del software LEXNET-UAEM 1.0.6.

Para generar las bases de datos que se ingresan al programa GEPHI, se debe ingresar primero la lista de vocabulario del infante al software

LEXNET-UAEM 1.0.6, y a partir de la información recabada del análisis de coocurrencia ejecutado anteriormente, el software selecciona las palabras contenidas en la lista de palabras del infante con sus relaciones correspondientes, arrojando diferentes archivos, uno de ellos con una lista de actores o nodos (en este caso palabras) que componen la red y otra con una lista de las relaciones o interacciones entre ellos, con información de la frecuencia de las coocurrencia entre las palabras.

Este proceso de obtención de base de datos por vocabulario individual ejecutado por el software LEXNET-UAEM 1.0.6 y la generación de redes léxicas con sus valores estadísticos, tarda aproximadamente 3 minutos, lo cual lo hace una herramienta rápida y fácil de usar por terapeutas, investigadores, maestros, etcétera.

### **Generación de palabras aprendibles**

Además de contar con la red individual de cada niño con sus valores estadísticos propios, al ingresar la lista con el vocabulario de cada niño al software LEXNET-UAEM 1.0.6, se obtiene una lista de palabras aprendibles, que se refieren a las palabras que no están presentes en el vocabulario de los niños. Para esto, al ingresar una lista conformada por las palabras vocabularios infantes obtenidos por el Inventario MacArthur-Bates del Desarrollo de Habilidades Comunicativas CDI con la versión del Inventario de Comprensión y Producción del Lenguaje en Infantes Mexicanos (ICPLIM),

LEXNET-UAEM 1.0.6 hace una búsqueda de las palabras que aún no están presentes en el vocabulario de cada infante y genera una lista de palabras aprendibles.

Posterior a la obtención de la lista de palabras aprendibles, el sistema puede generar otras listas de acuerdo a cualquiera de los 3 criterios o teorías de crecimiento de red que se mencionaron en el Capítulo II del presente trabajo, poniendo en primer lugar las palabras con mayor probabilidad de ser aprendidas eventualmente, de acuerdo a estos 3 criterios:

- Criterio 1: Preferential attachment model (Palabra nueva relacionada a palabra conocida con mayor indegree)

El programa ordenará las palabras aprendibles en función del valor de INDEGREE de las palabras conocidas (de mayor a menor). Una vez organizado con base en el nuevo orden, se seleccionarán las palabras que originalmente aparecían como TARGET y que no aparecen SOURCE, dejando solo las palabras aprendibles con mayor probabilidad de ser aprendidas de acuerdo a este criterio.

- Criterio 2: Prefferential acquisition model (Palabra nueva con mayor indegree)

El programa ordenará las palabras aprendibles tomando en cuenta valor de INDEGREE SOURCE más alto para ordenarlas de mayor a menor.

- Criterio 3: Lure of the associates (Palabra nueva con mayor indegree de palabras conocidas)

El programa deberá determinar el orden de la lista de las palabras aprendibles de acuerdo a cuantas veces las palabras de la lista de aprendibles aparecen como target en la lista de palabras conocidas.

## **2) Aplicación del sistema**

Para aplicar el sistema evaluamos el vocabulario de 10 niños cuyas edades oscilaban entre los 16 y 32 meses, y 10 vocabularios conformados con palabras aleatorias, generados por el sistema LEXNET-UAEM 1.0.6, igualando el número de palabras que conformaban el vocabulario de los 10 niños. Posteriormente se compararon pares de vocabularios que estuvieran conformados con el mismo número de palabras pero que se diferenciaban en la edad de los infantes. El procedimiento de la aplicación del sistema se describe en 8 pasos; 1. Selección del corpus para el análisis de coocurrencia, 2. Selección de palabras para la generación de valores de relación sintáctica-semántica entre palabras (palabras objetivo) para formar la red global, 3. Generación de red global (red modelo), 4. Relaciones de palabras de vocabularios individuales, 5. Generación de red individual por niño con sus valores estadísticos propios, 6. Generación de red aleatoria igualando cantidad de palabras dentro del vocabulario del niño, 7. Comparación

vocabularios misma cantidad pero diferente edad y 8. Generación de listas palabras aprendibles.

### **1. Selección del corpus para el análisis de coocurrencia**

Se utilizó una recopilación de diferentes corpora CHILDES, para el análisis de coocurrencia, sumando un total de 462,326 palabras. Para dicho análisis se seleccionó un tamaño de ventana de contingencia de 5, omitiendo dentro del conteo las palabras gramaticales.

### **2. Selección de palabras para la generación de valores de relación sintáctica semántica entre palabras (palabras objetivo) para formar la red global**

Como palabras objetivo se seleccionaron los sustantivos incluidos en el Inventario II MacArthur-Bates del desarrollo de habilidades Comunicativas (Jackson-Maldonado et al., 2005), incluyendo las 123 palabras que forman parte del Inventario de Comprensión y Producción del Lenguaje en Infantes Mexicanos (ICPLIM), elaborado por Hernández-Padilla y Alva (2011). De tales inventarios se seleccionaron los sustantivos que pertenecían a las categorías semánticas de animales, juguetes, alimentos, vehículos, cuerpo, ropa, artículos del hogar, exterior, muebles y personas, sumando un total de 445 sustantivos para el análisis de coocurrencia.

### 3. Generación de red global (red modelo)

Los datos obtenidos del análisis de coocurrencia con los 445 sustantivos, generaron bases de datos con información de relación entre palabras, tales bases de datos se ingresan en el programa de visualización de redes GEPHI para formar la red global (red modelo) con sus valores estadísticos propios. (ver fig. 3)

id	Label
1	abeja
2	Animal
3	Bicho
4	Borrego
5	Burro
6	Caballo
7	Caracol
8	Conejo
9	Elefante
10	Gallina
11	Gato
12	Gusano
13	Hormiga
14	Jirafa
15	León
16	Lobo
17	Mariposa
18	Mono
19	Mosca
20	Paloma
21	Pájaro
22	Panda
23	

Source	Target	Weight	Label	Type	IS	OS
1	13 Caballo	1	Relacion de	Directed		4
2	32 Mariposa	2	Relacion de	Directed		4
3	388 Casa	1	Relacion de	Directed		4
4	431 Niño	1	Relacion de	Directed		4
5	13 Caballo	2	Relacion de	Directed		8
6	22 Gallina	1	Relacion de	Directed		8
7	24 Gato	1	Relacion de	Directed		8
8	37 Pájaro	1	Relacion de	Directed		8
9	41 Perro	1	Relacion de	Directed		8
10	49 Ratón	2	Relacion de	Directed		8
11	54 Vaca	1	Relacion de	Directed		8
12	65 Carro	1	Relacion de	Directed		8
13	110 Jugo	1	Relacion de	Directed		8
14	388 Casa	1	Relacion de	Directed		8
15	415 Bebé	1	Relacion de	Directed		8
16	426 Mamá	1	Relacion de	Directed		8
17	1 abeja	2	Relacion de	Directed		5
18	19 Conejo	2	Relacion de	Directed		5
19	75 Tren	4	Relacion de	Directed		5
20	202 Cola	1	Relacion de	Directed		5
21	264 Cepillo	1	Relacion de	Directed		5
22	358 Flor	1	Relacion de	Directed		5
23						

Figura 3. Bases de datos generados por el software, nodos (derecha) y aristas (izquierda) para formar red global

### 4. Relaciones de palabras de vocabularios individuales

Posteriormente se ingresaron al software LEXNET UAEM 1.0.6, 10 vocabularios de infantes de entre 16 y 32 meses, para generar las bases de datos de las palabras que conforman sus vocabularios con sus valores de relación correspondientes. Los vocabularios de los 10 niños se obtuvieron mediante el Inventario II MacArthur-Bates del Desarrollo de Habilidades

Comunicativas, con la versión añadida del Inventario de Comprensión y Producción del Lenguaje en Infantes Mexicanos (ICPLIM). A diferencia de la versión original del Inventario II, en donde los padres solo tienen la opción de reportar palabras que sus hijos comprenden y dicen, en esta versión se consideran además las palabras que solamente comprenden y que todavía no expresan verbalmente.

### 5. Generación de red individual por niño con sus valores estadísticos propios

Las bases de datos generadas en el proceso anterior por el software LEXNET UAEM 1.0.6, serán ahora ingresadas en el programa de visualización de redes GEPHI, para obtener la red de cada niño con sus valores estadísticos propios (fig. 4)

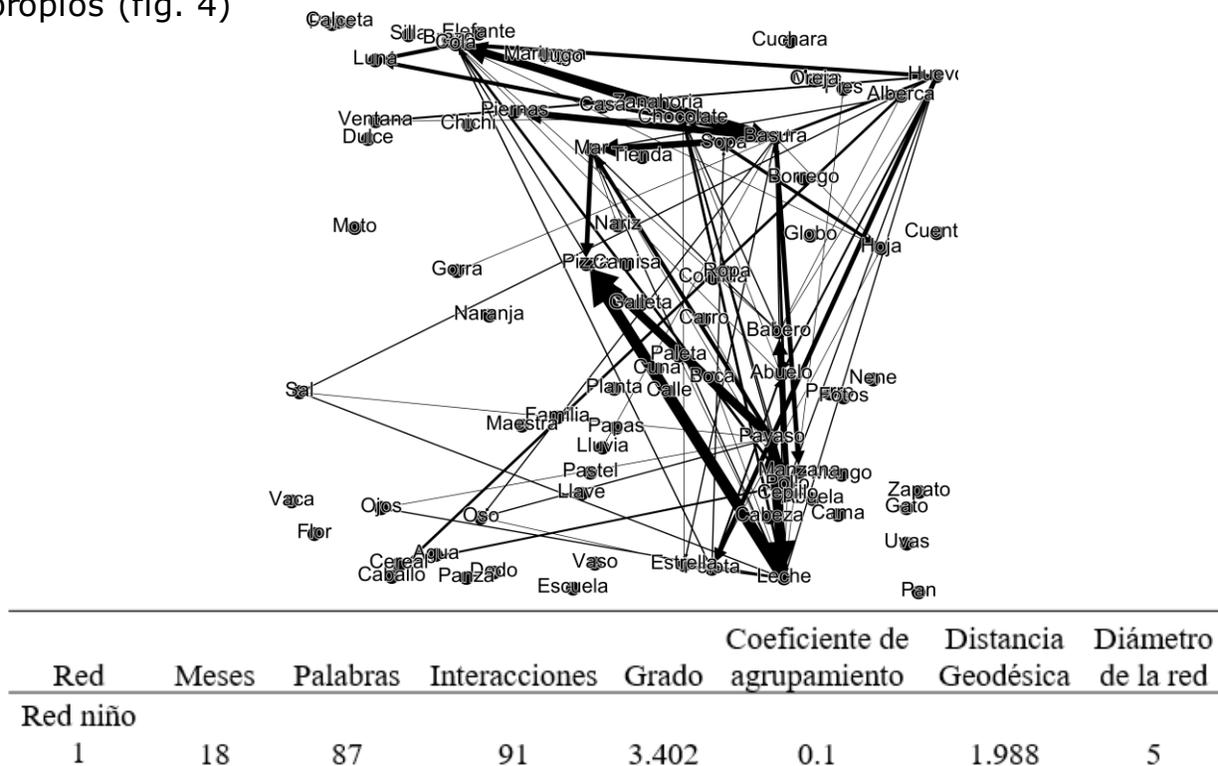


Figura 4. Red individual configurada a partir del vocabulario individual de un niño de 18 meses, con sus valores estadísticos

## **6. Generación de red aleatoria igualando cantidad de palabras dentro del vocabulario del niño**

En este paso, en el software LEXNET UAEM 1.0.6, se selecciona la opción de generar una red conformada con palabras aleatorias igualando la cantidad de palabras que compone el vocabulario individual de cada niño. Posteriormente las bases de datos generadas se ingresan en el programa de visualización de redes GEPHI, para obtener los valores estadísticos de la red aleatoria. Este paso tiene como objetivo probar la validez de los valores estructurales generados por el programa.

## **7. Comparación vocabularios misma cantidad pero diferente edad**

Se hizo una comparación entre pares de vocabularios, que estuvieran conformados con el mismo número de palabras pero que se diferenciaron en la edad de los infantes, por ejemplo, se hizo la comparación con dos vocabularios de 41 palabras, el primer vocabulario pertenecía a un niño de 19 meses y el otro a un niño de 25 meses, esto para explorar el peso potencial de cada uno de los valores estadísticos que se generan de la red.

## **8. Generación de listas palabras aprendibles**

Además de contar con la red individual de cada niño con sus valores estadísticos propios, con el sistema mostrado en el presente trabajo, al ingresar la lista con el vocabulario de cada niño al software LEXNET UAEM 1.0.6, se obtendrá una lista de palabras aprendibles, que se refieren a las

palabras que no están presentes en el vocabulario de los niños. Este proceso se divide en 3 pasos: 1. Definición de lista de vocabulario por niño, 2. Generación lista de palabras aprendibles y 3. Reordenamiento de palabras aprendibles según los 3 principios de crecimiento de red.

### ***1. Definición de lista de vocabulario por niño***

Se ingresa una lista conformada por las palabras vocabularios infantes obtenidos por el Inventario MacArthur-Bates del Desarrollo de Habilidades Comunicativas CDI con la versión del Inventario de Comprensión y Producción del Lenguaje en Infantes Mexicanos (ICPLIM).

### ***2. Generación lista de palabras aprendibles***

A partir de la lista de vocabulario ingresada anteriormente, se generará otra lista con las palabras que aún no están presentes en el vocabulario de cada infante. Es decir, la lista de palabras aprendibles estará compuesta por las palabras que aún no tiene adquiridas el infante al momento de la evaluación y que estén dentro del Inventario MacArthur-Bates del Desarrollo de Habilidades Comunicativas CDI con la versión del Inventario de Comprensión y Producción del Lenguaje en Infantes Mexicanos (ICPLIM).

### ***3. Reordenamiento de palabras aprendibles según los 3 principios de crecimiento de red***

El tercer paso, consta de generar 3 diferentes listas de palabras aprendibles, de acuerdo a los 3 principios de crecimiento red (palabra nueva relacionada a palabra conocida con mayor indegree, palabra nueva con mayor indegree y palabra nueva con mayor indegree de palabras conocidas); es decir, el programa organiza las palabras aprendibles de acuerdo a la probabilidad de ser aprendidas tomando en cuenta los criterios de cada principio de crecimiento de red.

## Resultados

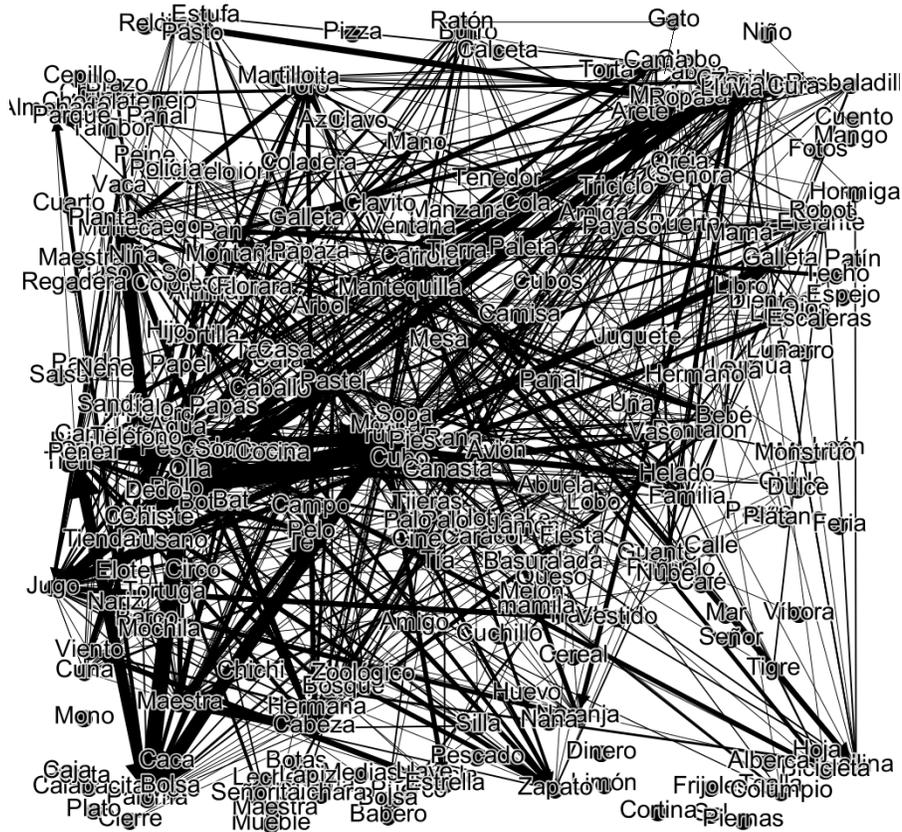
Los resultados que se presentan para análisis fueron obtenidos de un corpus conformado por 462,326 palabras, obtenidas de la base de datos CHILDES. Dicho corpus incluye el habla dirigida a infantes y el habla de infantes de conversaciones en las que participaban niños con edades de entre 10 meses a 12 años con sus padres o cuidadores. El corpus mencionado fue procesado mediante el programa LEXNET-UAEM 1.0.6 para obtener los valores de coocurrencia de las palabras objetivo incluidas en las listas que fueron definidas para el mismo programa. Estas listas inicialmente estaban compuestas por 445 sustantivos seleccionados del Inventario MacArthur-Bates del Desarrollo de Habilidades Comunicativas CDI con la versión del Inventario de Comprensión y Producción del Lenguaje en Infantes Mexicanos (ICPLIM), sin embargo la lista final con el reporte de las relaciones se redujo a 258 sustantivos que fueron aquellos que se encontraban tanto en las listas como en el corpus.

El análisis de coocurrencia con este corpus generó en primera instancia, las frecuencias de las relaciones entre palabras (ver fig. 5). Con dicha base de datos de frecuencia, se obtuvieron los valores estadísticos de lo que llamamos red global, la cual contenía los 257 sustantivos y sus relaciones correspondientes (ver fig. 6).

Simulación de Co-ocurrencia y Degrees -reporte global grande.csv

No.	I.S.	SOURCE	I.T.	TARGET	1o	2o	3o	4o	5o	FR-	SM-	FR+	SM+	FA	INDEGREE	OUTDEGREE	DEGREE
64	19	Conejo											81		22	20	42
65			4	Animal				1	1	0		2			2.47		
66			8	Bicho					1	0		1			1.23		
67			24	Gato	7	3	5	3	2	0		20			24.69		
68			29	Jirafa				1		0		1			1.23		
69			39	Pato				1		0		1			1.23		
70			41	Perró	3	3		1		0		7			8.64		
71			49	Ratón		1				0		1			1.23		
72			51	Tigre	1					0		1			1.23		
73			117	Manzana		1				0		1			1.23		
74			135	Pollo	3	6	2	1	2	0		14			17.20		
75			139	Sal	1	1	2		1	0		5			6.17		
76			155	Zanahoria	1	2		1	1	0		5			6.17		
77			239	Muñeca					1	0		1			1.23		
78			388	Casa	2	3		3		0		8			9.88		
79			413	Amiga		1		1		0		2			2.47		
80			415	Bebé		1			1	0		2			2.47		
81			425	Maestra			1	1		0		2			2.47		
82			426	Mamá	1	1			1	0		3			3.7		
83			442	Señora			1			0		1			1.23		
84			443	Señorita	1	1			1	0		3			3.7		

Fig. 5 Ejemplo de los datos de las frecuencias de relaciones entre palabras, generadas por el análisis de coocurrencias



Palabras	Interacciones	Grado	Coefficiente de agrupamiento	Distancia Geodésica	Diámetro de la red
258	691	6.174	0.105	2.95	7

Fig. 6 Red global con sus valores estadísticos

## **Resultados de análisis de vocabularios individuales**

Una vez que se contó con los valores de la red global como referencia para generar las redes léxicas individuales, se ingresaron al sistema LEXNET-UAEM 1.0.6 los vocabularios correspondiente a 10 niños con edades entre 16 y 32 meses ( $M=22.4$ ). De estos vocabularios se obtuvieron los valores estadísticos de la estructura de las redes léxicas, así como las redes gráficas para cada uno de los niños (fig. 7 a 16). Con el propósito de realizar una comparación y con ello una validación de los datos que se obtienen el sistema, se generaron los valores estadísticos de la red de 10 vocabularios conformados por palabras aleatorias. Los promedios de los valores estadísticos de los 10 vocabularios de los niños fueron comparados con los promedios de los valores estadísticos de las 10 redes generadas con palabras aleatorias. En general, los resultados de esta comparación mostraron que los valores estructurales de las redes de los niños reflejan una estructura más local, con una mayor cohesión entre palabras, lo que es indicador del efecto mundo pequeño. Las comparaciones entre los promedios de estos valores estadísticos, mostraron que los promedios de los vocabularios de los niños en el grado ( $M=2.753$ ;  $DE=1.892$ ) y el coeficiente de agrupamiento ( $M=0.085$ ;  $DE=.048$ ) fueron más altos que los promedios de las redes aleatorias en los mismos valores: grado ( $M=1.110$ ;  $DE=.979$ ) y coeficiente de agrupamiento ( $M=.029$ ;  $DE=.031$ ).

Estos promedios fueron sometidos a una prueba de ANOVA de una sola vía, cuyos resultados se muestran en la tabla 1.

ANOVA						
		Sum of Squares	df	Mean Square	F	Sig.
PALABRAS	Between Groups	.000	1	.000	.000	1.000
	Within Groups	32569.800	18	1809.433		
	Total	32569.800	19			
INTERACCIONES	Between Groups	5985.800	1	5985.800	.569	.460
	Within Groups	189338.000	18	10518.778		
	Total	195323.800	19			
GRADO MEDIO	Between Groups	13.487	1	13.487	5.942	.025
	Within Groups	40.859	18	2.270		
	Total	54.347	19			
COEFICIENTE DE AGRUPAMIENTO	Between Groups	.016	1	.016	9.700	.006
	Within Groups	.030	18	.002		
	Total	.046	19			
DISTANCIA GEODÉSICA	Between Groups	.778	1	.778	2.286	.148
	Within Groups	6.126	18	.340		
	Total	6.904	19			
DIÁMETRO DE LA RED	Between Groups	3.200	1	3.200	.865	.365
	Within Groups	66.600	18	3.700		
	Total	69.800	19			

Tabla 1. Resultados de los análisis estadísticos de la comparación de los valores estructurales de las redes de los niños y redes aleatorias. Los valores de Grado Medio y Coeficiente de agrupamiento mostraron diferencias significativas ( $p < .05$ ) entre grupo.

Las diferencias observadas en grado y coeficiente de agrupamiento entre los dos grupos de redes, sugieren que el instrumento es sensible a medir de forma válida la estructura de los vocabularios de los niños, estas diferencias entre redes aleatorias y redes de vocabularios corresponden a datos reportados en otros estudios (Steyvers & Tenenbaum, 2005). En tal estudio se encontró que altos puntajes en estos valores son característicos de una estructura del lenguaje y que se presentan en menor porcentaje en

redes de conformadas con palabras aleatorias, esto debido a que eliminan cualquier patrón de crecimiento ordenado.

Asimismo, los altos valores en coeficiente de agrupación y el alto promedio de grado de un nodo encontrados en las redes conformadas por los vocabularios de los niños, sugiere que algunos nodos pueden estar actuando como hubs o clusters. Estas características facilitan la interacción entre nodos y a su vez la adición de nuevas palabras, lo cual sugiere estar relacionado íntimamente con la estructura local de la red y eventualmente con el desarrollo lingüístico (Beckage et. al 2011). Tales características, coinciden con el efecto de redes de mundo pequeño. Dicho efecto permite a altas cantidades de estructura local combinarse con el acceso global de la red, y se sugiere están presentes en el lexicón de los adultos. Esto añade información a la hipótesis de que la estructura del lenguaje de los niños tiene características similares observadas a la de los adultos, sugiriendo que el léxico infantil puede ser una versión reducida del sistema de memoria semántica en adultos, en donde las representaciones semánticas de las primeras palabras aprendidas (edades tempranas) codifican relación semántica entre palabras, sugiriendo entonces que el sistema (lexicon) podría ser como el de un adulto, desde sus inicios.

Esto sugiere entonces, que al estar organizado como el sistema del lexicón adulto, el cual es un sistema de crecimiento auto-organizado y ciertas características de la red, permiten la interacción entre palabras y a la

vez su crecimiento, de igual manera; en el lexicon infantil ciertas características de la red van a permitir la interacción entre palabras y el crecimiento de la misma. Tal crecimiento puede llevarse a cabo entonces, siguiendo los principios de crecimiento empleados en otros estudios por diferentes autores (Hills, Maouene, Riordan, & Smith 2010), los cuales reportan que los nodos con muchos links (que se denominan nodos de alto grado) tienen más probabilidades de agregar nuevos vínculos que aquellos nodos con menos enlaces.

Es de llamar la atención que las diferencias observadas se hayan obtenido con una muestra tan pequeña el presente estudio (10 vocabularios por condición), sin embargo esta misma condición impide una mejor interpretación en aquellos valores en donde no se observaron diferencias significativas, como la distancia geodésica.

### **Resultados de comparación de vocabularios equivalentes**

Con el propósito de obtener resultados descriptivos que mostraran las diferencias entre vocabularios que contenían la misma cantidad de palabras pero que pertenecían a niños con edades distintas, se conformaron pares de vocabularios igualados en el número de palabras que conformaban tal vocabulario. Dichas comparaciones se muestran en las siguientes tablas y figuras.

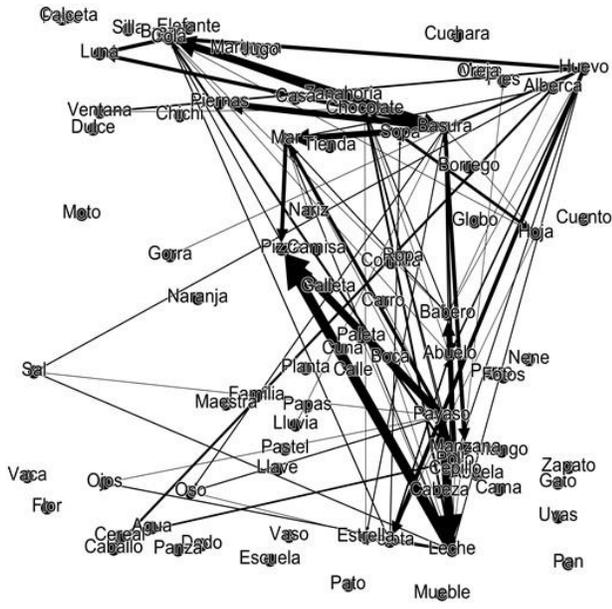


Figura 7. Red niño 1

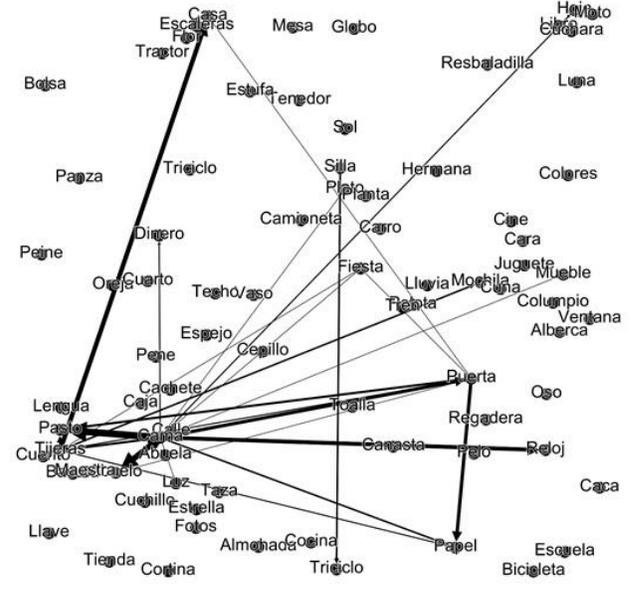


Figura 7. Red niño 2

Red	Meses	Palabras	Interacciones	Grado	Coefficiente de agrupamiento	Distancia Geodésica	Diámetro de la red
Red niño 1	18	82	91	3.402	0.1	1.988	5
Red niño 2	25	81	26	1.198	0.043	1.82	3

Tabla 2. Comparación de vocabularios equivalentes red niño 1 y red niño 2

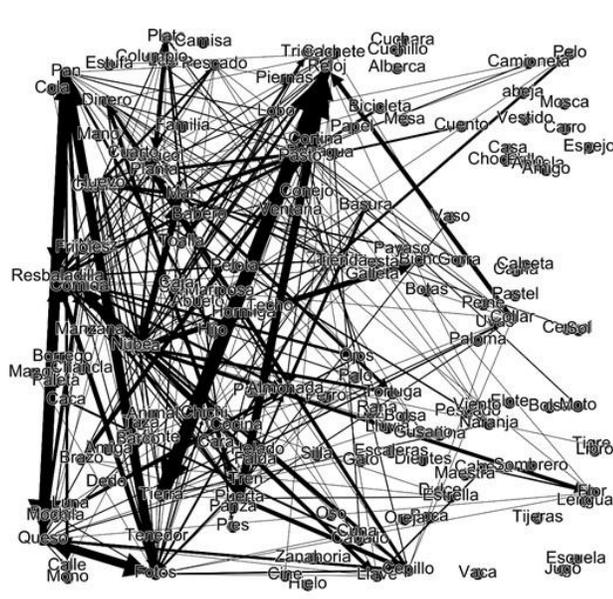


Figura 9. Red niño 3

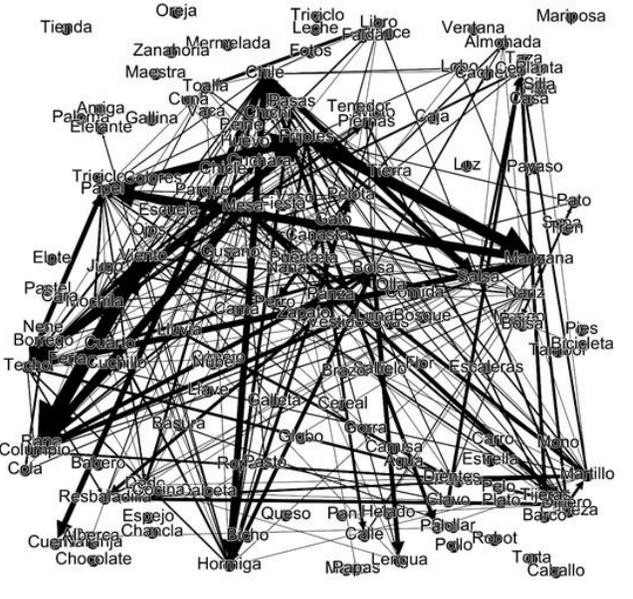


Figura 10. Red niño 4

Red	Meses	Palabras	Interacciones	Grado	Coefficiente de agrupamiento	Distancia Geodésica	Diámetro de la red
Red niño 3	19	158	321	5.785	0.104	2.68	6
Red niño 4	24	158	299	5.487	0.106	2.6	6

Tabla 3. Comparación de vocabularios equivalentes red niño 3 y red niño 4





pueden estar desempeñando un papel fundamental de conexión entre palabras. La función de estos clusters o palabras fungen como superconectores, mejorando la velocidad de procesamiento de la red, creando atajos que permiten que la recuperación de información se vea privilegiada (Borge-Holthoefer & Arenas 2010).

De igual manera que en los análisis anteriores, el tamaño de la muestra (10 vocabularios por condición), impide una mejor interpretación en aquellos valores en donde no se observaron diferencias significativas, como el caso la distancia geodésica que resultó similar en ambos grupos, tanto en el de vocabularios de edades tempranas como en el de redes de edades mayores.

Redes	Meses	Palabras	Interacciones	Grado	Coefficiente de agrupamiento	Distancia Geodésica	Diámetro de la red
Chicos	18.6	83.4	116.2	3.394	0.1014	2.0502	4.6
Grandes	26.2	83.4	84.2	2.1122	0.0704	2.0294	4

Tabla 7. Comparación de promedios redes niños de edades tempranas y redes de niños de redes mayores

## Conclusiones

En primera instancia con la creación del programa, se reafirma la necesidad del trabajo interdisciplinario, debido a que es indispensable la colaboración de diversas y diferentes disciplinas y, en general, la colaboración de especialistas procedentes de diversas áreas como lo son profesionales en psicología, terapia de lenguaje y profesionales de la programación computacional, esto para lograr un producto de tal calidad que pueda cubrir los requerimientos necesarios, para que como en este caso; sea posteriormente utilizado tanto como un instrumento de evaluación e intervención, así como de investigación.

El programa LEXNET-UAEM 1.0.6, el cual fue desarrollado de manera interdisciplinaria, ofrece la posibilidad de evaluar y observar el desarrollo lingüístico mediante las redes léxicas. Tal programa brinda una nueva manera de analizar y apreciar las características del desarrollo de los infantes, ya que no solo da a conocer el número y tipo de palabras que el infante tiene adquiridas, sino también, permite conocer la relación y organización semántica de las mismas lo que ofrece información valiosa para la estimulación y potencialización del aprendizaje de palabras y a su vez, prevenir futuras dificultades en su desarrollo. A la par, el programa LEXNET-UAEM 1.0.6 podrá ayudar a contestar preguntas en la investigación que nos permitan entender cómo se organiza el vocabulario de las personas, en

diferentes poblaciones y cuáles son los factores cognitivos, ambientales, entre otros, que influyen en dicha organización. Es decir, el presente programa, eventualmente brindará la posibilidad de aportar información en cuanto a qué factores están siendo determinantes para la organización del léxico de los infantes, permitiendo examinar diferentes aspectos o criterios. Entre ellos, ofrece la oportunidad de analizar diferentes tipos de palabras, utilizando diferentes corpus, incluso centrando la atención en el contexto del ambiente aprendizaje que rodea al menor, analizando el ambiente de aprendizaje del lenguaje y observar las relaciones existentes entre la organización de los vocabularios. Por lo tanto, el presente trabajo, es un primer acercamiento para conocer más sobre los diferentes procesos e implicaciones en la organización de las palabras en el léxico mental de los niños, lo que parece relevante para conocer cómo se está llevando a cabo este proceso y qué relación e impacto tiene sobre el desarrollo lingüístico.

Los resultados obtenidos por el programa LEXNET-UAEM 1.0.6, añaden información a la hipótesis de que la estructura del lenguaje de los niños tiene características similares observadas a la de los adultos, el cual es un sistema de crecimiento auto-organizado y ciertas características de la red, permiten la interacción entre palabras y a la vez su crecimiento. Esto tomando en cuenta los altos valores en coeficiente de agrupación y el alto promedio de grado de un nodo encontrados en las redes conformadas por los vocabularios de los niños, en contraste a los valores obtenidos por las redes aleatorias.

Tales resultados sugieren que algunos nodos pueden estar actuando como hubs o clusters. Estas características facilitan la interacción entre nodos y a su vez la adición de nuevas palabras, lo cual sugiere estar relacionado íntimamente con la estructura local de la red y eventualmente con el desarrollo lingüístico (Beckage et. al 2011).

Sin embargo, en los resultados obtenidos en el presente estudio, a pesar de que se observan diferencias de un grupo a otro en algunos valores; que en estos casos fueron el promedio de grado, el coeficiente de agrupamiento y las interacciones; no se pueden sacar conclusiones determinantes debido al tamaño de la muestra empleada en los análisis. Esto da pie a realizar análisis con una muestra más grande y con una diferencia entre edades mayores a la muestra empleada en el presente estudio, para poder identificar y determinar el peso potencial de los valores estadísticos que se miden con el programa y poder reafirmar de manera contundente las hipótesis referidas en otros estudios. Sin embargo, los objetivos de crear el sistema LEXNET-UAEM 1.0.6 abarcaron además, proporcionar información relevante para la estimulación del desarrollo lingüístico. Por lo que el presente nuevo método de evaluación no pretende etiquetar o diagnosticar los retrasos de los infantes, si no que uno de los usos más relevantes es conocer y analizar los procesos de adquisición de palabras desde edades tempranas para poder proporcionar a todos por igual, los recursos necesarios para crear una situación de aprendizaje óptima y enriquecedora, de acuerdo

a sus características y particularidades de cada infante; esto partiendo de las palabras que ya tienen adquiridas utilizándolas como facilitador del aprendizaje de nuevas palabras.

Es decir, el análisis de la red individual de los infantes además de brindar información acerca de la organización de la misma y sus valores estructurales correspondientes, proporcionará eventualmente, información relevante para poder llevar a cabo una estimulación lingüística personalizada para lograr el aprendizaje de nuevas palabras, las cuales serán facilitadas debido a que estarán relacionadas semánticamente con las que ya conoce.

Hacer este tipo de análisis del desarrollo lingüístico plantea la importancia de considerar no solo las deficiencias o debilidades del niño, sino, analizar otros factores que toman en cuenta las fortalezas o habilidades, en conjunto con el contexto del ambiente de aprendizaje. Lo que supone aceptar la diversidad como elemento de progreso y de riqueza de la colectividad, y por otro lado, asumir las diferencias en las personas como elementos que van a enriquecer su aprendizaje y el de los demás (Melero, 2004). Por lo que el presente estudio toma en cuenta el principio de que cada niño/a tiene características, intereses, capacidades y necesidades de aprendizaje distintos y deben ser los sistemas educativos, educadores, padres de familia y herramientas de evaluación e intervención los que deberían de estar diseñados teniendo en cuenta la amplia diversidad de dichas características y necesidades.

## Referencias

- Aldana, M. (2006). Redes complejas. *Recuperado a partir de <http://www.fis.unam.mx/~max/English/notasredes.pdf>*. Recuperado a partir de <http://www.fis.unam.mx/~max/Spanish/notasredes.pdf>
- Auza, A. (2009). ¿Qué es el trastorno del lenguaje? Un acercamiento teórico y clínico a su definición. *Lenguaje*, 37(2), 365-391.
- Añorve, N. (2015) Adquisición de palabras en infantes por medio de las redes léxicas, (Tesis inédita de Maestría). Universidad Autónoma del Estado de Morelos, Facultad de Comunicación Humana, Morelos, México.
- Bates E, Goodman JC. (1997) On the inseparability of grammar and the lexicon: Evidence from acquisition, aphasia, and real-time processing. *Language and Cognitive Processes* 1997;12(5-6):507-584.
- Aldana, M. (2006). Redes complejas. *Recuperado a partir de <http://www.fis.unam.mx/~max/English/notasredes.pdf>*. Recuperado a partir de <http://www.fis.unam.mx/~max/Spanish/notasredes.pdf>
- Beckage, N., Smith, L., & Hills, T. (2011). Small worlds and semantic network growth in typical and late talkers. *PloS one*, 6(5), e19348.
- Beckage, N., Aguilar, A., & Colunga, E. (2015). Modeling lexical acquisition through networks. En *Proceedings of CogSci*. Recuperado a partir de <https://tsel.cs.colorado.edu/~nibe3229/papers/BeckageColungaCogSci2015.pdf>
- Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., & Hwang, D. (2006). Complex networks: Structure and dynamics. *Physics Reports*, 424(4-5), 175-308. <https://doi.org/10.1016/j.physrep.2005.10.009>
- Borovsky, A., Ellis, E. M., Evans, J. L., & Elman, J. L. (2016). Semantic Structure in Vocabulary Knowledge Interacts With Lexical and Sentence Processing in Infancy. *Child Development*, 87(6), 1893-1908. <https://doi.org/10.1111/cdev.12554>

- Bloom, P. (1993). *Language Acquisition: Core Readings*. Estados Unidos: Harvester Wheatsheaf.
- Cancho, R. F. i., & Sole, R. V. (2001). The small world of human language. *Proceedings of the Royal Society B: Biological Sciences*, 268(1482), 2261–2265. <https://doi.org/10.1098/rspb.2001.1800>
- Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological review*, 82(6), 407.
- Colunga, E., & Sims, C. E. (2011). Early talkers and late talkers know nouns that license different word learning biases. En *Proc. of CogSci* (Vol. 11). Recuperado a partir de <http://palm.mindmodeling.org/cogsci2011/papers/0598/paper0598.pdf>
- Fisher, C., Hall, D. G., Rakowitz, S., & Gleitman, L. (1994). When it is better to receive than to give: Syntactic and conceptual constraints on vocabulary growth. *Lingua*, 92, 333–375.
- Fodor, J. A. (1983). *The modularity of mind: An essay on faculty psychology*. MIT press. Recuperado a partir de [http://books.google.com/books?hl=en&lr=&id=e7nrSeibJZYC&oi=fnd&pg=PP9&dq=%22characteristics+of+each+of+the+presumed+faculties,+then%22+%22are+psychological+faculties+from+a+particular%22+%22the+faculty/modularity+issues+from+what+I%27ll+call%22+%22though+I+think+the+revival+of+the+faculty+psychology+program%22+&ots=ooYFVySB7\\_&sig=msMDFmKyND1km2mqRQugSO09q8A](http://books.google.com/books?hl=en&lr=&id=e7nrSeibJZYC&oi=fnd&pg=PP9&dq=%22characteristics+of+each+of+the+presumed+faculties,+then%22+%22are+psychological+faculties+from+a+particular%22+%22the+faculty/modularity+issues+from+what+I%27ll+call%22+%22though+I+think+the+revival+of+the+faculty+psychology+program%22+&ots=ooYFVySB7_&sig=msMDFmKyND1km2mqRQugSO09q8A)
- Gasser, M., & Smith, L. B. (1998). Learning nouns and adjectives: A connectionist account. *Language and cognitive processes*, 13(2–3), 269–306.
- Hills, T. T., Maouene, J., Riordan, B., & Smith, L. B. (2010a). The associative structure of language: Contextual diversity in early word learning. *Journal of Memory and Language*, 63(3), 259–273. <https://doi.org/10.1016/j.jml.2010.06.002>

- Hills, T. T., Maouene, J., Riordan, B., & Smith, L. B. (2010b). The associative structure of language: Contextual diversity in early word learning. *Journal of Memory and Language, 63*(3), 259–273.  
<https://doi.org/10.1016/j.jml.2010.06.002>
- Jackson-Maldonado, D. (2004). Spanish-speaking Mexican later-talkers: Vocabulary and gestures. *Anuario de Psicología/The UB Journal of Psychology, 35*(2), 257.
- Luce, P.D., Pisoni, D.B., and Goldinger, S.D. 1990. Similarity neighborhoods of spoken words. In G. Altmann (ed.), *Cognitive models of speech perception: Psycholinguistic and computational perspectives*. Cambridge, MA: MIT Press, 122-147.
- Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and hearing, 19*(1), 1.
- McRae, K., Cree, G. S., Seidenberg, M. S., & McNorgan, C. (2005). Semantic feature production norms for a large set of living and nonliving things. *Behavior research methods, 37*(4), 547–559.
- Melero, M. L. (2004). *Construyendo una escuela sin exclusiones: una forma de trabajar en el aula con proyectos de investigación*. Ediciones Aljibe. Recuperado a partir de  
<https://books.google.com.mx/books?id=sxmUAAAACAAJ>
- Murillo, E., & Belinchón, M. (2013). Patrones comunicativos multimodales en la transición a las primeras palabras: cambios en la coordinación de gestos y vocalizaciones. *Infancia y Aprendizaje, 36*(4), 473–487.  
<https://doi.org/10.1174/021037013808200258>
- Murphy, M. L. (2003). *Semantic relations and the lexicon: antonymy, synonymy, and other paradigms*. Cambridge, UK ; New York: Cambridge University Press.

- Nematzadeh, A., Fazly, A., & Stevenson, S. (2014). Structural differences in the semantic networks of simulated word learners. *CogSci Proceedings*, 1072–1077.
- Newman, M. E. (2003). The structure and function of complex networks. *SIAM review*, 45(2), 167–256.
- Pablo, M. N. (2003). Adquisición del lenguaje. El principio de la comunicación. *CAUCE, Revista de Filología y su Didáctica*, (26), 321–347.
- Paul, R., Looney, S. S., & Dahm, P. S. (1991). Communication and socialization skills at ages 2 and 3 in late-talking young children. *Journal of Speech, Language, and Hearing Research*, 34(4), 858–865.
- Perianes-Rodríguez, A., Olmeda-Gómez, C., & de Moya-Anegón, F. (2008). Introducción al análisis de redes. *El Profesional de la Informacion*, 17(6), 664–669. <https://doi.org/10.3145/epi.2008.nov.10>
- Rescorla, L., Mirak, J., & Singh, L. (2000). Vocabulary growth in late talkers: lexical development from 2;0 to 3;0. *Journal of Child Language*, 27(2), 293–311. <https://doi.org/10.1017/S030500090000413X>
- Steyvers M, Tenenbaum JB (2005) The large-scale structure of semantic networks: statistical analyses and a model of semantic growth. *Cognitive Science* 29: 41–78.
- Styles, S., & Plunkett, K. (2011). Early links in the early lexicon: Semantically related word-pairs prime picture looking in the second year. *Lexical representation: A multi-disciplinary approach*, 51–88.
- Traxler, M. J., & Gernsbacher, M. A. (Eds.). (2006). *Handbook of psycholinguistics* (2nd ed). Boston: Elsevier.
- Vitevitch, M. S. (2008). What Can Graph Theory Tell Us About Word Learning and Lexical Retrieval? *Journal of Speech Language and Hearing Research*, 51(2), 408. [https://doi.org/10.1044/1092-4388\(2008/030\)](https://doi.org/10.1044/1092-4388(2008/030))

- Vivas, J., & Vivas, J. (2009). Modelos de memoria semántica. *J. Vivas (Comp.) Evaluación de redes semánticas. Instrumentos y Aplicaciones. MdP: Eudem*. Recuperado a partir de <http://psicoblogs.com.ar/archivos/Psicologia%20Cognitiva/eb7e60Modelos%20de%20memoria%20semantica.pdf>
- Weismer, S. E., Tomblin, J. B., Zhang, X., Buckwalter, P., Chynoweth, J. G., & Jones, M. (2000). Nonword repetition performance in school-age children with and without language impairment. *Journal of Speech, Language, and Hearing Research, 43*(4), 865–878.
- Weldon, A. (31 Marzo 2014). Language Nutrition for All. [Mensaje en un blog]. Recuperado de [http://gafcp.org/communicate/tmp\\_blog/BlogLanguageNutrition#.VHjmhDGG-Sr](http://gafcp.org/communicate/tmp_blog/BlogLanguageNutrition#.VHjmhDGG-Sr)
- Werker, J. F., & Curtin, S. (2005). PRIMIR: A developmental framework of infant speech processing. *Language learning and development, 1*(2), 197–234.

## Anexos

### Desarrollo del software (Descripción técnica)

1. Descripción de las funciones del software (análisis de la coocurrencia de las palabras y valores estadísticos obtenidos para crear redes léxicas)
2. Simulación del funcionamiento del software
3. Generación de la red con los datos obtenidos por el software y sus valores estadísticos propios por medio de un programa de visualización (Gephi)

1. Descripción de las funciones del software (análisis de la coocurrencia de las palabras y valores estadísticos obtenidos para crear redes léxicas)

Durante esta fase, se seleccionaron y describieron las funciones características del software. El programa permitirá realizar un análisis de relación entre palabras generando valores cuantitativos sobre las magnitudes de dichas relaciones y sobre la configuración de las redes léxicas que se generan. Las cuales se explican en diferentes funciones.

### Funciones

#### Función 1: Análisis de coocurrencias

Evaluar la relación entre palabras dentro de un texto, por medio de un análisis de las coocurrencias observadas entre dichas palabras. Tal análisis permite generar una base de datos con los valores requeridos para la creación de una red léxica (*red global*)

Además, otra opción del programa será trabajar con una base de datos global ya preestablecida

Descripción: El programa funciona con el input de uno o varios archivos de texto, sobre los que se realizará el análisis de coocurrencia, generando valores sobre la relación entre palabras dentro de ese texto, es decir, se contabilizará cuántas veces dos palabras coocurren dentro de un texto en un rango de palabras determinado. Al utilizar las estadísticas de coocurrencia como el índice de relación semántica, en las redes, la palabra A (Palabra SOURCE) se conectará a la palabra B (Palabra TARGET) si aparece dentro de las primeras cinco palabras (o el tamaño de ventana que se elija) que siguieron a la palabra B, produciendo una red dirigida en el que una relación de A B no implica B también está conectado a A.

El término Palabra SOURCE se otorgará a la palabra que se esté tomando como referencia para contabilizar las palabras que aparecen o coocurren después de esta, y la Palabra TARGET será esa palabra que está después de una palabra SOURCE.

El tamaño de la ventana se refiere al rango de palabras dentro de las cuales se contabilizará la coocurrencia de una palabra con otra. Y el tamaño de la ventana de cinco se basó en el trabajo previo que investigó el poder predictivo a través de múltiples tamaños de ventana (Hills et al., 2010b).

Las palabras sobre las que se generarán los valores de las relaciones estarán definidas en dos listas, entendiéndose que se generará un valor de relación entre cada una de las palabras de la Lista 1 con cada una de las palabras de la Lista 2

#### Función 2: Filtro por vocabulario individual

De la base de datos global obtenida anteriormente o de la base de datos global preestablecida, se filtrarán a una nueva base de datos solo las palabras que pertenecen a la lista del vocabulario de un infante determinado al cual llamaremos *palabras conocidas* y a otro archivo (base de datos) se

filtrarán las palabras restantes al cual llamaremos *palabras aprendibles*. Esto para diferenciar cuáles palabras no están en el vocabulario de los infantes, y reordenarlas en función a los diferentes criterios de crecimiento de la red que se explicará en funciones posteriores.

### Función 3: Generación de sets de aleatorias

A partir de la base de datos global obtenida por medio del análisis de coocurrencia (o de la base de datos preestablecida), se generarán sets de diferentes números de palabras con sus relaciones y frecuencias correspondientes, esto con el objetivo de generar redes aleatorias con el mismo número de palabras pero con diferentes palabras elegidas aleatoriamente

Se podrá seleccionar un número cualquiera de palabras (1-2000) y el número de sets o bases de datos que se requieren (1-1000), y el programa seleccionará ese número en palabras con sus relaciones y frecuencias, arrojando el número de sets seleccionado en diferentes bases de datos. Estos sets de palabras aleatorias, servirán para formar redes que no tengan una estructura propia del lenguaje, a diferencia de las obtenidas de los vocabulario de los niños, y al realizar una comparación entre ellas se espera observar un efecto mayor de redes de mundo pequeño en las redes de vocabularios de los infantes en comparación con las aleatorias, tal y como lo observaron otros autores como Beckage et al., (2011).

### Función 4: Reordenamiento palabras aprendibles por criterio

Esta función será realizada con la información individual de cada infante; las palabras conocidas y las palabras aprendibles. La función 4 reordenará la lista de las palabras aprendibles poniendo en primer lugar las palabras con mayor probabilidad de ser aprendidas eventualmente, de acuerdo a 3 criterios:

-Criterio 1: a palabra conocida con mayor degree (Preferential attachment model)

El programa ordenará las palabras aprendibles en función del valor de INDEGREE de las palabras conocidas (de mayor a menor). Una vez organizado con base en el nuevo orden, se seleccionarán las palabras que originalmente aparecían como TARGET y que no aparecen SOURCE, dejando solo las palabras aprendibles con mayor probabilidad de ser aprendidas de acuerdo a este criterio.

-Criterio 2: aprendibles con mayor degree (Prefferential acquisition model)

El programa ordenará las palabras aprendibles tomando en cuenta el valor de INDEGREE SOURCE más alto para ordenarlas de mayor a menor.

-Criterio 3: a más palabras que conozca (Lure of the associates)

El programa deberá determinar el orden de la lista de las palabras aprendibles de acuerdo a cuantas veces las palabras de la lista de aprendibles aparecen como target en la lista de palabras conocidas.

2. Simulación del funcionamiento del software, se han realizado manualmente las funciones que deberá ejecutar el software. En primera instancia, se seleccionó un corpus lingüístico, y se eligió una lista de tipos de palabras: sustantivos. Para realizar el análisis de coocurrencia dentro del corpus, se localizan las palabras que pertenecen a la lista 1 y 2 (que contienen las mismas palabras) y se contabilizan las 5 primeras palabras que aparecen después de cada una. En este conteo efectuado por el programa, se genera un grado de relación que permitirá formar una base de datos con valores de fuerza de asociación de una palabra con otra, se contará también con el número de palabras con las que coocurren un mayor número de veces. Esta información servirá para crear la red modelo (que es la red generada con todas las palabras de la lista) y la red individual (la cual

conectará las palabras del vocabulario de cada niño de acuerdo a las estadísticas de coocurrencia de las palabras dentro del corpus seleccionado).

Para la red modelo y para la red individual, se promediaron tres estadísticas de la red sobre todas las palabras en la red. Estos son indegree, coeficiente de agrupamiento, y la distancia geodésica (*in-degree, clustering coefficient, and geodesic distance*).

El Indegree ( $k_i$ ) se calcula como el número de tipos de palabra única precedente de una palabra dada en el corpus; es decir, el número de aristas en la red que apuntan hacia esa palabra. Con base al análisis de coocurrencia, el indegree será el número de veces que una palabra aparece o coocurre después de otra.

El coeficiente de agrupación de un nodo ( $c_i$ ) se calculará mediante la determinación de cuántas conexiones existen entre los vecinos más cercanos de ese nodo (nodo  $i$ ), es decir, el coeficiente de agrupación es entonces la fracción de conexiones observadas, ( $l_i$ ), entre los que sea posible. Este valor se promedia en todos los nodos para conseguir un coeficiente medio de agrupación para toda la red. El coeficiente de agrupación ( $c_i$ ) es una medida común de la estructura local dentro de la red y ha relacionado con la velocidad de procesamiento, así como el reconocimiento y la recuperación de información.

La distancia geodésica se calcula como la longitud media de camino más corto entre dos nodos, y es una medida de acceso global de las palabras en la red. En conjunto, estas tres estadísticas proporcionan tres diferentes niveles de información sobre la estructura de la red, lo cual permitirá a su vez, conocer las particularidades del repertorio del léxico de cada niño.

Como ya se comentó anteriormente, para construir las redes individuales, el software conectará las palabras del vocabulario de cada niño de acuerdo a las estadísticas de coocurrencia de las palabras en un ambiente

de aprendizaje de diferentes corpora, como el CHILDES corpus, -una gran base de datos de aproximadamente de 2 millones de palabras del habla dirigida a infantes-, como también de la base de datos de Normas de Asociación de Palabras (NAP) para el español de México, entre otras. De este modo, los nodos de la red individual serán las palabras conocidas por ese niño, y los enlaces entre nodos reflejan las relaciones semánticas en el lenguaje en general. Las redes resultantes se verán limitadas en tamaño y contenido a las palabras formuladas dentro del cuestionario.

Los vocabularios de cada infante, que servirán para formar las redes léxicas, serán recolectados a través de un reporte parental ampliamente utilizado, el CDI Inventario MacArthur-Bates, una medida fiable y válida de desarrollo léxico, (Dale, 1996 cit. en Beckage et al. 2011) lo que brindará grandes ventajas en cuanto validez y confiabilidad puesto que la aplicación final resulta igual de económica e eficiente al ser la fuente principal un CDI.

### 3. Generación de la red con los datos obtenidos por el software y sus valores estadísticos propios por medio de un programa de visualización de redes llamado Gephi

Una vez obtenidas las bases de datos con la información de relaciones entre palabras, se ingresarán de manera individual (cada base de datos) al programa de visualización de redes llamada Gephi, dicho programa se encuentra disponible en línea de manera gratuita. Tal programa generará gráficamente la red con sus valores estadísticos propios.