

Manual de uso

LEXPRO



VNIVERSIDAD
D SALAMANCA



UNIVERSITAS
Miguel Hernández



MINISTERIO
DE CIENCIA
E INNOVACION



AGENCIA
ESTATAL DE
INVESTIGACIÓN

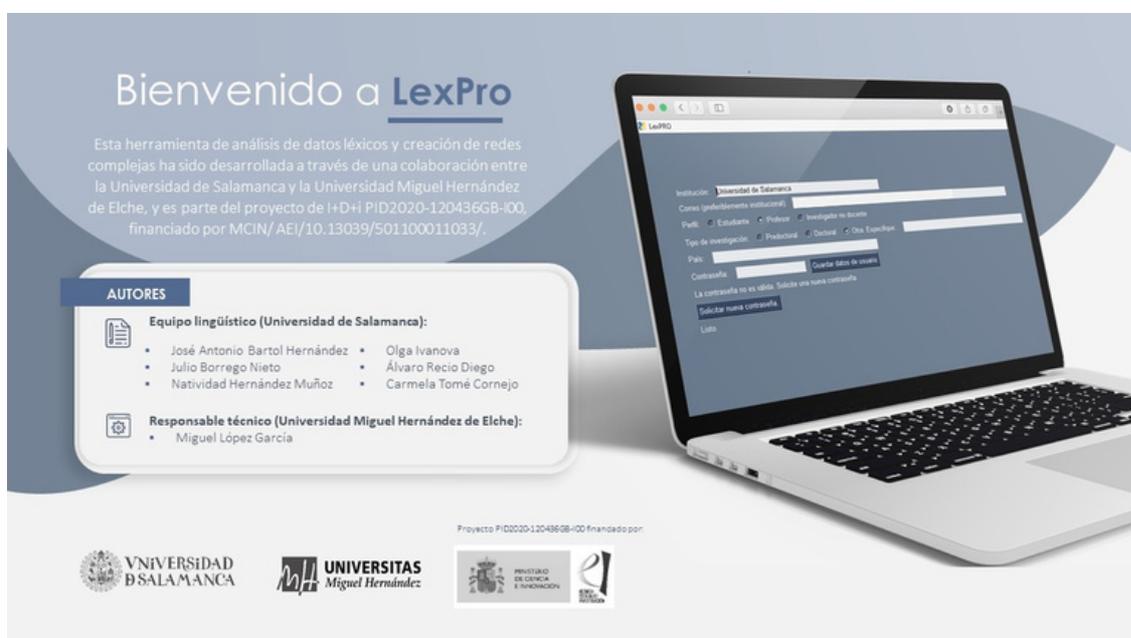
Proyecto PID2020-120436GB-I00 financiado por:

ÍNDICE

- 01** Para comenzar a utilizar LexPRO
- 02** Para preparar el fichero de datos
- 03** Para obtener los datos de disponibilidad léxica
- 04** Para generar grafos y analizar redes léxicas
- 05** Para exportar los datos

PARA COMENZAR A UTILIZAR LEXPRO

1. Descargue en su dispositivo la herramienta de la página web <https://dispogram.usal.es/>.
2. Abra la carpeta de archivos y ejecute LexPRO- Aplicación.
3. Se abrirá una ventana de bienvenida con los datos de la herramienta. Señale INICIAR LEXPRO.



4. Complete los datos de usuario y solicite el envío de la contraseña a su correo electrónico. La contraseña tiene una validez de un mes. Cuando la contraseña caduque puede solicitar otra nueva.

5. A continuación, cargue el archivo de datos que quiere analizar en SELECCIONAR FICHERO DE DATOS. En la parte inferior de la pantalla puede seguir el proceso de carga de datos del archivo.

El fichero de datos debe estar en formato EXCEL de acuerdo con las pautas que se describen en el apartado siguiente. Puede encontrar una plantilla del modelo de Excel en <https://dispogram.usal.es/>.

6. En la pestaña CARGA puede comprobar que su fichero se ha incorporado correctamente a la herramienta. En ella se indica el número de centros de interés, el número de informantes en cada uno de ellos y las variables de agrupación de los informantes.



Nº	Nombre	NP	SEXO	CENTRO	AMBITO	RESIDENCIA	NIVEL SOCIOCULTURAL
1	Partes_Cuerpo	600	HOMBRE	PUBLICO	RURAL	RURAL	BAJO
2	Ropa	600	MUJER	PRIVADO	URBANO	URBANA	MEDIOBAJO
3	Muebles	600					MEDIOALTO
4	Partes_casa	600					ALTO
5	Ciudad	600					
6	Campo	600					
7	Transporte	600					
8	Trabajo_Campo	600					
9	Profesiones	600					
10	Animales	600					

PARA PREPARAR EL FICHERO DE DATOS

7. El documento de Excel constará de varias pestañas. El nombre de las pestañas debe aparecer con mayúscula inicial (el resto en minúscula) y las etiquetas formadas por más de una palabra se unen mediante barra baja (Partes_Cuerpo). La primera pestaña se denominará "Informantes"; la segunda, "Variables", y a continuación se nombrarán tantas pestañas como categorías o centros de interés se hayan utilizado.

Informantes | Variables | Partes_Cuerpo | Ropa | Muebles

8. En la primera pestaña, INFORMANTES, se añaden los datos de las variables de los informantes. En la primera columna (CODIGO_INFORMANTE), debe indicarse el número de cada informante a partir de la segunda fila, según se muestra en el ejemplo. En la segunda columna y sucesivas, se recogen, codificadas con números, sus características individuales a partir de la fila 2, puesto que el nombre de las variables, que debe aparecer en mayúsculas (SEXO, CENTRO, etc.), se incluye en la fila 1. Es obligatorio que haya al menos una variable de descripción o agrupación de los participantes.

	A	B	C	D	E	F	G
1	CODIGO_INF	SEXO	CENTRO	ÁMBITO	RESIDENCIA	NIVEL SOCIOCULTURAL	
2	1	1	2	2	2	3	
3	2	1	2	2	2	2	
4	3	1	2	2	1	2	
5	4	1	2	2	2	2	
6	5	2	2	2	2	2	
7	6	1	2	2	2	3	
8	7	1	2	2	2	4	
9	8	1	2	2	1	2	
10	9	2	2	2	2	2	
11	10	1	2	2	2	2	
12	11	1	2	2	2	3	
13	12	2	2	2	1	1	
14	13	2	2	2	1	1	
15	14	1	2	2	1	3	
16	15	1	2	2	1	3	
17	16	1	2	2	1	2	
18	17	1	2	2	1	2	
19	18	2	2	2	2	2	
20	19	1	2	2	2	3	

9. En la segunda pestaña, VARIABLES, se especifican las opciones de cada una de las variables de descripción de los informantes que contengan los datos. Hay que situarlas en el mismo orden de columna en el que se han codificado en la pestaña de los informantes. La fila 1 contiene las denominaciones de las variables (SEXO) y en las siguientes se especifican sus opciones en orden creciente según su codificación: HOMBRE (codificada como 1) y MUJER (2). Es obligatorio que haya al menos una variable de descripción o agrupamiento de los participantes. En esta pestaña aparecen en mayúsculas todos los términos.

	A	B	C	D	E	F
1	SEXO	CENTRO	ÁMBITO	RESIDENCIA	NIVEL SOCIOCULTURAL	
2	HOMBRE	PÚBLICO	RURAL	RURAL	BAJO	
3	MUJER	PRIVADO	URBANO	URBANA	MEDIOBAJO	
4					MEDIOALTO	
5					ALTO	
6						
7						

10. En la tercera pestaña y siguientes se introducen los datos de los centros de interés, uno por pestaña; por tanto, habrá tantas pestañas como centros de interés. El nombre de la pestaña reflejará el nombre del centro de interés que utilice la herramienta (recuerde que esta denominación no puede contener espacios y debe llevar mayúscula inicial). La primera

fila de estas hojas debe dejarse libre para identificar correctamente a cada informante con sus respuestas. Cada una de las palabras debe ir en una casilla diferente.

	A	B	C	D	E	F	G	H	I	J
1										
2	cuello	pie/s	espalda	columna/s (l	dedo/s	mano/s	hombro/s	Barbiella	pierna/s	tronco/s
3	cerebro	mano/s	diente/s	pierna/s	frontal	temporal	parietal	occipital	cerebelo	fémur
4	hueso/s	uña/s	dedo/s	pie/s	mano/s	pecho	hombro/s	cara	ojo/s	boca
5	diente/s	páncreas	bazo	cerebro	cerebelo	dermis	epidermis	intestino/s	estómago	esófago
6	cabeza	brazo/s	muñeca/s	dedo/s	nariz/es	ojo/s	oreja/s	hueso/s	músculo/s	tendón/es
7	cabeza	oreja/s	pestaña/s	pecas	diente/s	caries	cuello	hueso/s	brazo/s	uña/s
8	brazo/s	dedo/s	mano/s	oreja/s	nariz/es	cabello/s	pecho	abdomen	antebrazo/s	pene
9	ojo/s	testículo/s	ombigo	dedo/s	uña/s	ceja/s	boca	articulación/	muñeca/s	menisco
10	corporales	cabeza	mano/s	pie/s	pelo/s	uña/s	hueso/s	músculo/s	órganos	células
11	pierna/s	ojo/s	oreja/s	pelo/s	labio/s	diente/s	lengua	piel	hueso/s	cerebro
12	ojo/s	costilla/s	dedo/s	lengua	corazón	pulmón/es	estómago	esófago	riñón/s	hígado
13	brazo/s	cuello	pie/s	estómago	páncreas	costilla/s	espalda	columna/s (l	cerebro	codo/s
14	cabeza	culo	pecho	pie/s	omoplato	hombro/s	muslo/s	cintura	tibia	peroné

Para EXPORTAR LOS DATOS DESDE WORD O TXT, pueden seguirse los pasos que se describen a continuación.

A

Si los datos de cada centro de interés se han transcrito en un archivo Word o txt con las respuestas de cada informante dispuestas en filas y separadas por comas, abra un documento de Excel y seleccione la pestaña de “Datos” > “Desde el texto/CSV”. A continuación, seleccione el archivo de texto correspondiente > “Importar”, y, en el cuadro de diálogo del asistente para la importación, marque “Coma” como delimitador y “Transformar datos” > “Cargar”. Si sus datos incluyen la codificación de los informantes, como en la imagen siguiente, puede añadir una coma tras tales dígitos para que no se exporten en la misma columna que la primera respuesta de los informantes. En Word, desde “Inicio”, seleccione “Reemplazar > Más > Especial > Cualquier número > Espacio”. Tal y como se muestra en la imagen, en el recuadro “Buscar:” aparecerá ^#, los caracteres comodín de “Cualquier número”. No olvide incluir un espacio tras ellos. A continuación, en “Reemplazar con:”, escriba una coma y pulse “Reemplazar todos”. Deberá repetir este proceso para cada centro de interés.

```
11111 001 01 boca, nariz, pene, intestino, hígado, esófago, tráquea, riñón, oreja, ojo, ano, pulmón, corazón,
11111 002 01 cara, estómago, pie, mano, nariz, hígado, intestino, rodilla, dedo, corazón, pulmón, oreja, codo,
11111 003 01 ojo, cabeza, pelo, nariz, boca, brazo, pierna, pecho, panza, pie, corazón, pulmón, faringe, dedo,
```

Buscar:

Reemplazar con:

B

Ya en Excel, es importante comprobar que no hay celdas en blanco entre las respuestas de un mismo informante y que las celdas con las respuestas no contienen espacios antes o después de estas. Para esto último, puede pasar una macro que elimine los espacios al inicio y al final de cada celda. Active la pestaña del programador si no está visible: “Archivo > Opciones > Personalizar cinta de opciones > Programador / Desarrollador”. A continuación, en esa pestaña, seleccione la opción “Macros”, en “Nombre de la Macro” escriba EliminarInicioFinal y pulse “Crear”. Al hacerlo se abre una ventana de Visual Basic. Borre el código que aparece en ella y copie este:

```
Sub EliminarInicioFinal()  
Dim celda As Range  
For Each celda In Selection  
celda.Value = LTrim(celda.Value)  
celda.Value = RTrim(celda.Value)  
Next  
End Sub
```

C

Cierre la ventana del Visual Basic, seleccione las celdas correspondientes (Ctrl+E) y ejecute la macro creada en cada hoja de respuestas: “Ctrl+E > Programador > Macros > EliminarInicioFinal > Ejecutar”.

PARA OBTENER LOS DATOS DE DISPONIBILIDAD LÉXICA

11. Una vez preparado el fichero de datos en Excel, inicie el programa y cargue el archivo (puntos 1-6 de este manual). A continuación, en la pestaña GRUPOS se pueden establecer los grupos de análisis. Por defecto aparece creado el grupo TODOS, que incluye a todos los participantes en el estudio. Para crear un nuevo grupo, se debe marcar la casilla CREAR GRUPO, añadir el nombre y seleccionar las variables de agrupación. El número de informantes de dicho grupo aparecerá automáticamente a su derecha. Puede regresar a esta pantalla en el momento que lo requiera para crear grupos de estudio.

GRUPOS	SEXO		CENTRO		ÁMBITO		RESIDENCIA		NIVEL SOCIOCULTURAL			
	HOMBRE	MUJER	PÚBLICO	PRIVADO	RURAL	URBANO	RURAL	URBANA	BAJO	MEDIABAJO	MEDIOALTO	ALTO
Todos	580	X	<input checked="" type="checkbox"/>									
Hombres	290	X	<input checked="" type="checkbox"/>									
Mujeres	390	X	<input checked="" type="checkbox"/>									
NIS_bajo	179	X	<input checked="" type="checkbox"/>									
NIS_medioabajo	319	X	<input checked="" type="checkbox"/>									
NIS_medioalto	123	X	<input checked="" type="checkbox"/>									
NIS_alto	59	X	<input checked="" type="checkbox"/>									

12. En la pestaña GLOBAL se obtienen automáticamente los datos cuantitativos generales de los datos incorporados en el fichero: TOTAL DE PALABRAS (*tokens*), VOCABLOS (*types* o palabras diferentes), PROMEDIO POR INFORMANTE (*token average*, el total de palabras dividido entre el número de informantes) e ÍNDICE DE COHESIÓN. Este indicador mide el grado de coincidencia de las respuestas entre los sujetos para un mismo centro de interés y se obtiene dividiendo el número de palabras por sujeto entre los vocablos (si el promedio de respuestas es de 20 y el centro de interés tiene 20 palabras diferentes, el índice de cohesión será 1). En la parte inferior de la pantalla se puede seguir el proceso de cálculo de cada uno de los centros de interés.

13. En la pestaña INFORMANTES se calcula de manera automática el número de palabras dadas por cada informante en cada uno de los centros de interés.

14. En la pestaña DISPONIBILIDAD se calculan los datos para cada uno de los centros de interés: DISPONIBILIDAD LÉXICA –fórmula de López-Chavez y Strassburger (2000)[1]–, DISPONIBILIDAD POR 100, FRECUENCIA RELATIVA (porcentaje que supone el vocablo en el total de las palabras/*tokens* de la muestra: si *perro* aparece mencionada 20 veces en las 100 palabras totales para Animales, obtendrá un 40% de frecuencia relativa), APARICIÓN (porcentaje de los informantes que producen una palabra: si *perro* es dicha por 82 informantes de 100 que participan en la muestra, tendrá un 82% de aparición) y FRECUENCIA ACUMULADA (suma de las frecuencias relativas de los términos con menor disponibilidad). La herramienta permite ordenar la obtención de datos por disponibilidad, por aparición y por frecuencia acumulada.

15. La pestaña COMPARACIÓN permite realizar comparaciones por centro de interés entre los grupos de informantes establecidos.



Seleccione el centro de interés que quiere comparar y los grupos entre los que quiere establecer la comparación. Para realizar las comparaciones el programa calcula los

[1] López-Chavez, J. y C. Strassburger (2000). “El diseño de una fórmula matemática para obtener un índice de disponibilidad léxica confiable”, *Anuario de Letras*, 38, 227-251.

listados de disponibilidad de todos los grupos seleccionados. La temporalización del proceso puede seguirse, como en las anteriores ocasiones, en la parte inferior de la pantalla.

B

Si no se indica lo contrario, las comparaciones se realizarán entre el conjunto de vocablos que suman los grupos comparados (el programa indica la cifra total de vocablos/palabras diferentes). Para establecer filtros se puede seleccionar un número mínimo de VOCABLOS POR DISPONIBILIDAD (por ejemplo, los 100 con mayor disponibilidad) o excluir aquellos que no alcanzan un determinado porcentaje de aparición en UMBRAL DE APARICIÓN, tanto en el listado general como en el listado de cada grupo.

C

Cada vez que se realiza un cambio en las condiciones de comparación –grupos, centro de interés o filtros– debe marcarse el botón ANALIZAR para que el programa vuelva a calcular los datos.

D

En el primer recuadro aparecen los datos cuantitativos de las comparaciones: NÚMERO DE INFORMANTES, PALABRAS TOTALES, VOCABLOS, NÚMERO DE PALABRAS EXCLUSIVAS aportadas por cada grupo y los porcentajes que suponen los vocablos compartidos y los vocablos excluidos en el conjunto de los vocablos dados por cada grupo. La suma de estos dos porcentajes no alcanza el 100% cuando se analizan más de dos grupos, porque habría que sumarle el tanto por ciento que suponen las palabras compartidas por dos o más grupos, pero no por todos. El número de palabras compartidas, que es el mismo para todos los grupos, aparece en el recuadro central junto con el índice de compatibilidad (el tanto por ciento que suponen las palabras compartidas si consideramos el total de vocablos de ambos conjuntos; por ejemplo, si hombres y mujeres tuvieran 40 vocablos cada uno y compartieran esos 40 vocablos, el grado de compatibilidad sería 100%, mientras que si los hombres tienen 40 vocablos y las mujeres otros 40 y solo comparten 20, la compatibilidad sería 33,33%).

GRUPO	Nº INFORMANTES	PALABRAS (TOKENS)	VOCABLOS (TYPES)	Nº EXCLUSIVAS	% EXCLUSIVAS	% COMPARTIDOS
España	80	1829	178	46	25,84%	38,76%
Cuba	80	1355	140	24	17,14%	49,29%
Perú	80	1668	134	17	12,69%	51,49%
Venezuela	80	1234	114	18	15,79%	60,53%

Nº COMPARTIDOS	SUMA DE CONJUN	COMPATIBILIDAD
69	252	27,38%

Las imagen previa y la siguiente presentan la comparación del centro de interés de *Animales* en cuatro capitales de países hispanohablantes (Bartol Hernández, 2022)[2]. Para eliminar las palabras dadas por un solo informante se ha realizado un corte en el 1,26 de aparición, pues la muestra de cada grupo es de 80 informantes.

E

En los recuadros del final de la página aparecen los VOCABLOS COMPARTIDOS por los grupos en contraste, ordenados por su índice de disponibilidad general en el conjunto, y las palabras aportadas solo por un grupo (VOCABLOS EXCLUSIVOS), las cuales se ordenan en función de la disponibilidad que alcanzan en su correspondiente listado.

Vocablos compartidos		Vocablos exclusivos			
Nº	PALABRA	España	Cuba	Perú	Venezuela
1	perro/a	1 buitre	tocororo	llama	chiguire
2	gato/a	2 ñu	jicotea	cocuyo	puercoespin
3	león/a	3 periquito	cotorra	chancho	cochino
4	tigre	4 maniquita	manatí	cóndor	zancudo
5	elefante	5 golondrina	zunzún	vicuña	zamuro
6	mono/a	6 lince	jutía	ganso	morrocoy
7	jirafa	7 salmón	almíquí	pulga	cunaguaro
8	caballo	8 piraña	puerco	otorongo	bachaco
9	vaca	9 orca	cunel	ñandú	mantarraya
10	cocodrilo	10 pez espada	alacrán	pericote	pez martillo

PARA GENERAR GRAFOS Y ANALIZAR REDES LÉXICAS

16. Como novedad con respecto a programas anteriores, la pestaña de GRAFOS incorpora la herramienta para el cálculo de los índices matemáticos de redes léxicas y la generación de gráficos. Asume como base teórica que el hecho de que dos palabras se asocien en la misma cadena de producción implica una proximidad en el lexicón mental. Un NODO es una palabra que presenta conexiones con otros nodos o palabras, una ARISTA es la línea imaginaria que une dos nodos y posee un PESO que representa la relación de proximidad de ambas palabras en las cadenas de producción.

Para comenzar, una vez señalado el centro de interés (SELECCIONAR CENTRO DE INTERÉS) y el grupo de informantes sobre el que se va a generar el gráfico (SELECCIONAR GRUPOS), se ha de indicar el tipo de gráfico que va a generarse según los datos de proximidad incluidos en él. La herramienta permite crear cuatro tipos de gráficos según se contabilice dicho grado de proximidad:

A

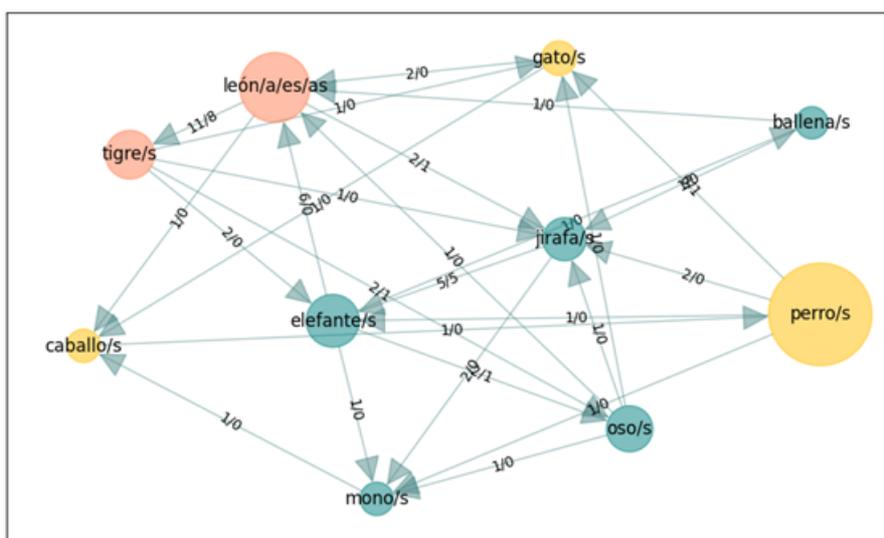
BIGRAM DIRIGIDO. El gráfico representa todas las parejas de palabras (*bigrams* o binomios) que aparecen en las respuestas teniendo en cuenta la dirección de producción. En la cadena *perro, gato, caballo, vaca* se contabilizarían tres *bigrams*: perro-gato, gato-caballo, caballo-vaca. Cada una de estas parejas generaría el valor

[2] Bartol Hernández, J. A. (2022). "Léxico panhispánico: animales", en C. Díaz Alayón (ed.), *Studia Philologica in honorem José Antonio Samper*. Madrid: Arco Libros/Academia Canaria de la Lengua, 97-134.

de 1 en la matriz de datos. El gráfico representa la matriz generada por el conjunto de todos los datos seleccionados:

	perro	gato	caballo	vaca
perro		1		
gato			1	
caballo				1
vaca				

En la imagen siguiente aparece un *bigram* dirigido para las 10 palabras más disponibles de *Animales*. Tanto este como los grafos siguientes se han elaborado a partir de los datos de Hernández Muñoz (2006)[3]. En el *bigram* dirigido cada una de las aristas incluye el peso de ambas direcciones y la flecha marca la dirección más frecuente: en el ejemplo, la flecha indica la dirección *león-tigre*, porque 11 participantes produjeron esta pareja y solo 8 *tigre-león*. En esta ocasión se ha igualado el ancho de las aristas a 1, pero las opciones de visualización permiten que el peso se refleje en su anchura. Igualmente, se ha mantenido como opción de visualización el tamaño de los nodos (con diferentes tamaños según su disponibilidad) y los *hubs* o comunidades coloreados (opción COLOREAR COMUNIDADES).

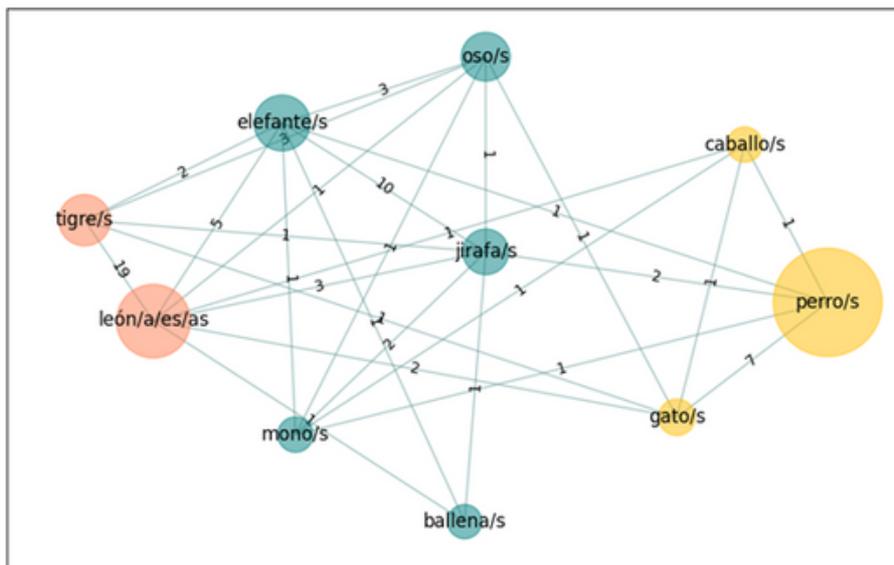


B

BIGRAM NO DIRIGIDO. El gráfico representa todas las parejas de palabras (*bigrams* o binomios) sin considerar la dirección en la cadena de producción.

[3] Hernández Muñoz, N. (2006). *Hacia una teoría cognitiva integrada de la disponibilidad léxica: el léxico disponible de los estudiantes castellano-manchegos*. Salamanca: Ediciones Universidad de Salamanca.

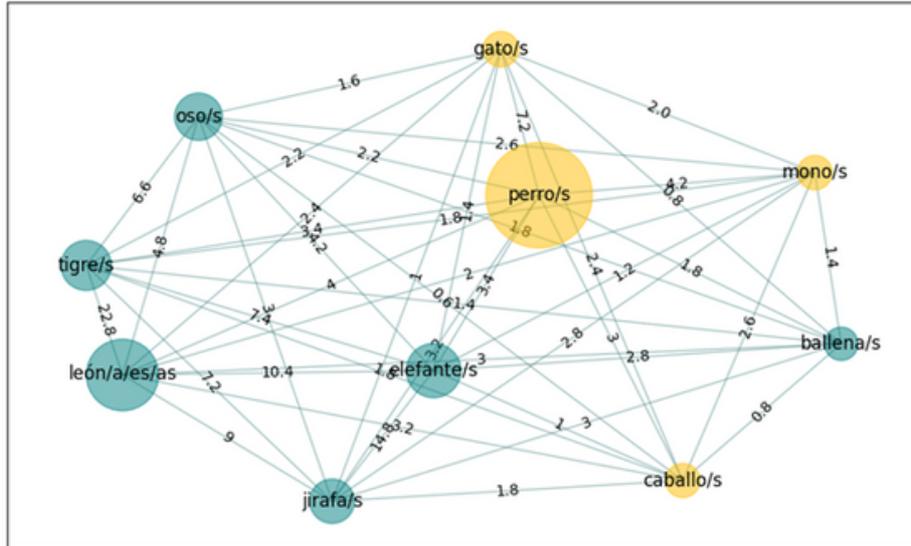
El ejemplo siguiente muestra el mismo conjunto de datos que el anterior (*bigram* dirigido para las 10 palabras más disponibles de *Animales*) sin especificar la direccionalidad. El número que aparece en las aristas es la frecuencia total de aparición, la suma de ambas direccionalidades. Así, la arista que une *tigre* y *león* presenta ahora un peso de 19.



LINEAL. Contabiliza todas las parejas de palabras ponderando la distancia a la que están hasta cinco posiciones. En la cadena *perro, gato, caballo, vaca*, cada una de estas parejas generaría diferentes valores en la matriz de datos según la distancia a la que se encuentran. Los pesos decrecen de forma lineal desde 1 (palabras adyacentes o distancia = 1) hasta 0,2 (distancia = 5). El resto de distancias tienen un peso de 0, de modo que se obtendría la siguiente matriz de datos según la cadena de producción propuesta.

	perro	gato	caballo	vaca
perro		1	0,8	0,6
gato	1		1	0,8
caballo	0,8	1		1
vaca	0,6	0,8	1	

Este sería el resultado del gráfico lineal para las 10 palabras más disponibles de *Animales*:



D

EXCEL. La herramienta permite incorporar un Excel donde se definen los criterios matemáticos de creación del gráfico dirigido o no dirigido. Para incorporar el fichero hay que pulsar el botón SELECCIONAR FICHERO EXCEL. Este fichero debe especificar las distancias entre palabras que se van a considerar en la columna A y los pesos asignados a cada distancia en la columna B. En el ejemplo de la imagen, si dos palabras se registran consecutivamente (por ejemplo, *perro-gato*) recibirán un peso de 1. Si aparece una palabra entre ambas (distancia 2), como en *perro-ratón-gato*, su peso será de 0,9, si aparecen dos (distancia 3), como en *perro-hámster-ratón-gato*, se le asignará un peso de 0,8, y así sucesivamente hasta una distancia máxima de 10. Ahora bien, la tabla de Excel no tiene por qué mantener la linealidad. Se puede proponer una reducción exponencial o incluso dar el mismo peso a varias distancias, por ejemplo, peso 1 a las distancias de 1, 2 y 3, y asignar peso 0 al resto.

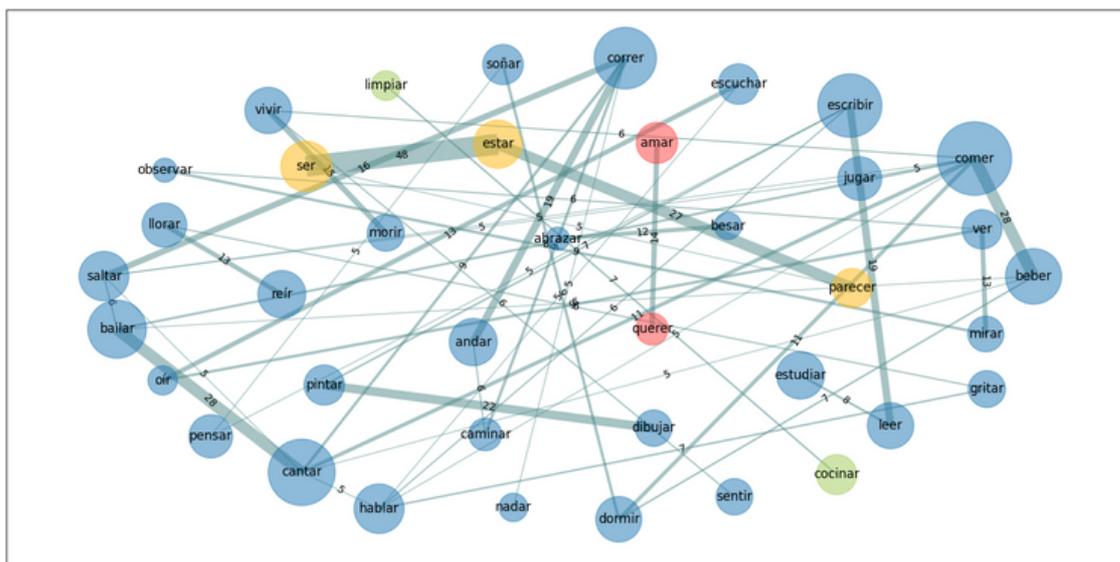
	A	B	C
1	1	1	
2	2	0,9	
3	3	0,8	
4	4	0,7	
5	5	0,6	
6	6	0,5	
7	7	0,4	
8	8	0,3	
9	9	0,2	
10	10	0,1	

17. Para generar un grafo debe seleccionarse la pestaña CREAR GRAFO. El gráfico aparecerá en una ventana nueva que puede expandirse. Las opciones de visualización

permiten seleccionar diferentes parámetros de diseño del grafo: TAMAÑO DE LOS NODOS (según su disponibilidad), LONGITUD DE ARISTA (para representar la proximidad de los nodos), RIGIDEZ DE ARISTA (rigidez del "muelle" que une cada palabra: a las aristas de mayor peso se les asignará una rigidez en la parte alta del rango, esto es, tendrán una mayor tendencia a mantener su longitud deseada que las aristas de menor peso), ANCHO DE ARISTA (para reflejar el peso de la asociación) y REPULSIÓN ENTRE NODOS CONECTADOS (para mejorar la visibilidad del gráfico y facilitar la identificación de las comunidades o *hubs*). Asimismo, se ofrece la posibilidad de colorear las comunidades que se distinguen en el grafo.

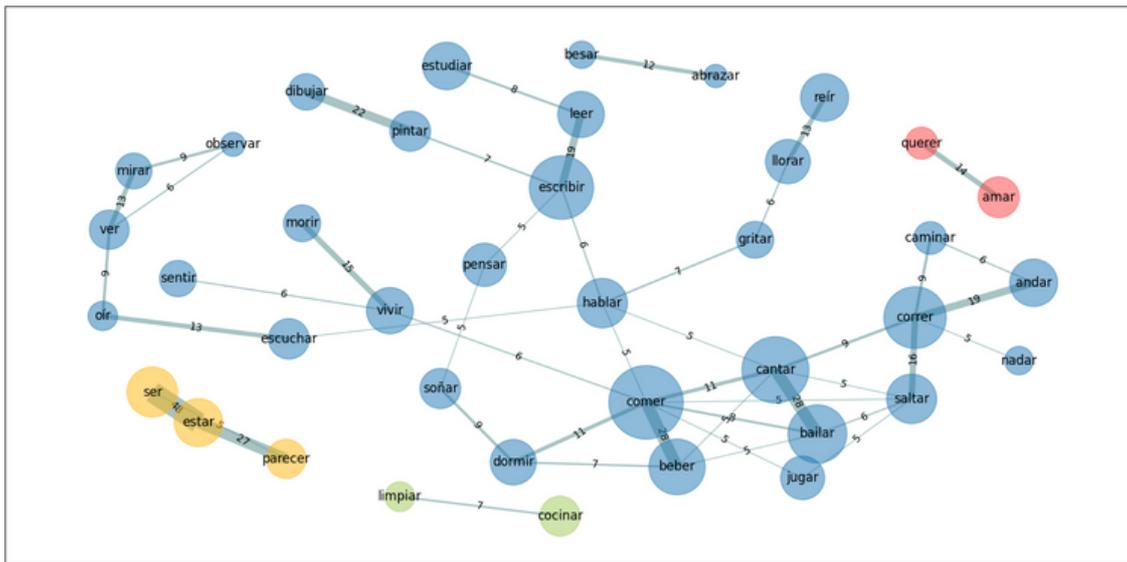


Las siguientes imágenes muestran la utilidad de la opción de repulsión entre nodos no conectados a la hora de identificar los *hubs* de un grafo. Las tres han sido creadas a partir del mismo conjunto de datos del centro de interés de *Verbos* (Tomé y Recio, 2022)[4] para las 50 palabras más disponibles, con la poda de 4 aristas y las modificaciones de repulsión de 40, 150 y 180. El gráfico de repulsión 180 ofrece una presentación más limpia y permite identificar unos *hubs* más concretos.

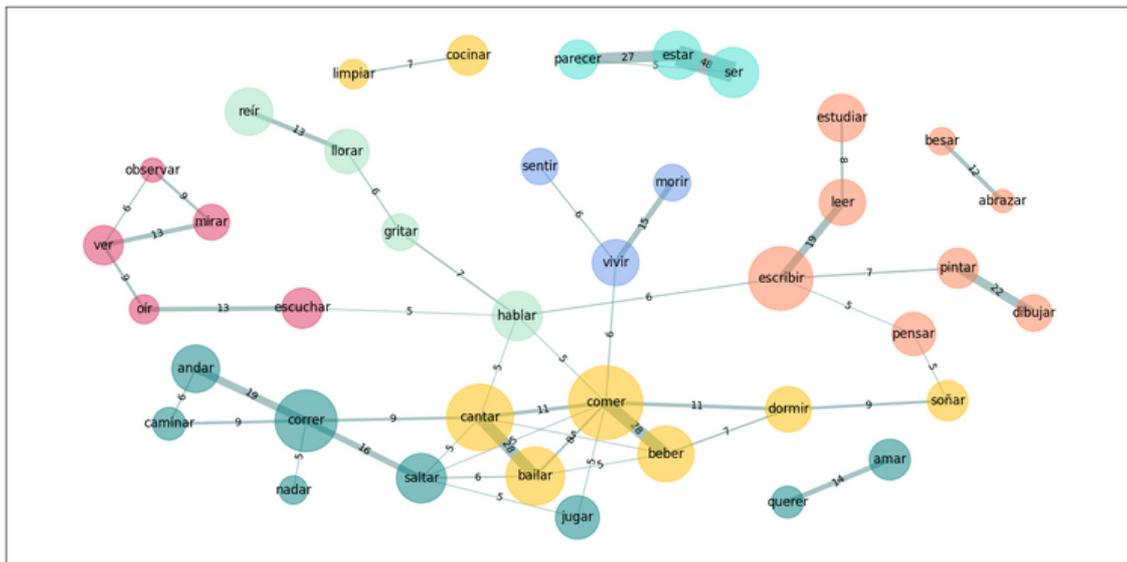


Repulsión entre nodos de 40

[4] Tomé Cornejo, C. y Recio Diego, Á. (2022). "De la disponibilidad léxica a la disponibilidad gramatical. Desarrollos metodológicos y aplicación didáctica", *TEJUELO. Didáctica de la lengua y la literatura*, 35(3), 299-336.



Repulsión entre nodos de 150



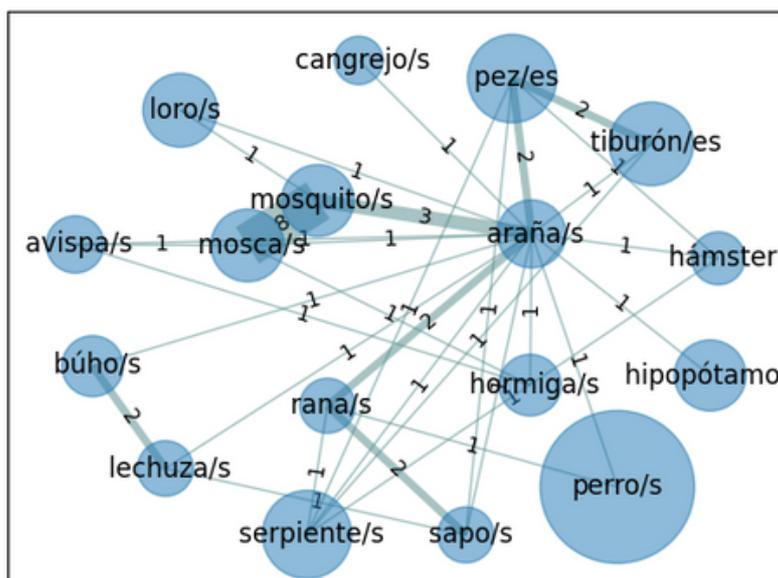
Repulsión entre nodos de 180

18. Se puede realizar un podado de nodos y de aristas para estudiar redes más sencillas, seleccionar los datos más significativos o facilitar la visualización. Para ello, en la sección SELECCIONAR FILTRADO POR NODOS Y ARISTAS se pueden indicar los nodos o aristas que se quieran excluir del gráfico. En el filtrado de nodos se pueden utilizar los criterios de índice de disponibilidad (VOCABLOS POR DISPONIBILIDAD), porcentaje de aparición (UMBRAL DE APARICIÓN) o número mínimo de nodos a los que tiene que estar conectado un determinado nodo (NODOS POR GRADO, si se selecciona 3, significa que se eliminarán aquellos nodos que no tengan al menos 3 asociados). Hay que tener en cuenta que el filtrado consiste en identificar los nodos con 3 o más asociados y mantener en el grafo a estos nodos y sus asociados. Los asociados no tienen por qué cumplir la condición.

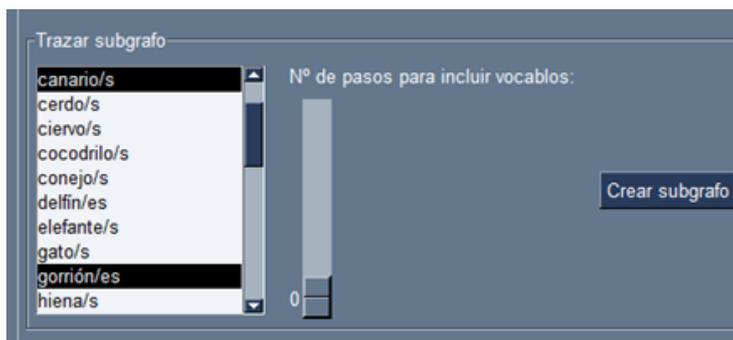
En el grafo aparecerán con distinto tono los nodos que pasan el filtro y los nodos que simplemente son asociados. En cuanto al filtrado por aristas, se pueden excluir las relaciones entre dos palabras que tengan menos de un determinado peso, tal y como este se haya definido en la configuración del tipo de grafo. Se omitirán las aristas con un peso igual o menor al elegido. Si un nodo no tiene aristas que cumplan el filtro, no será representado.

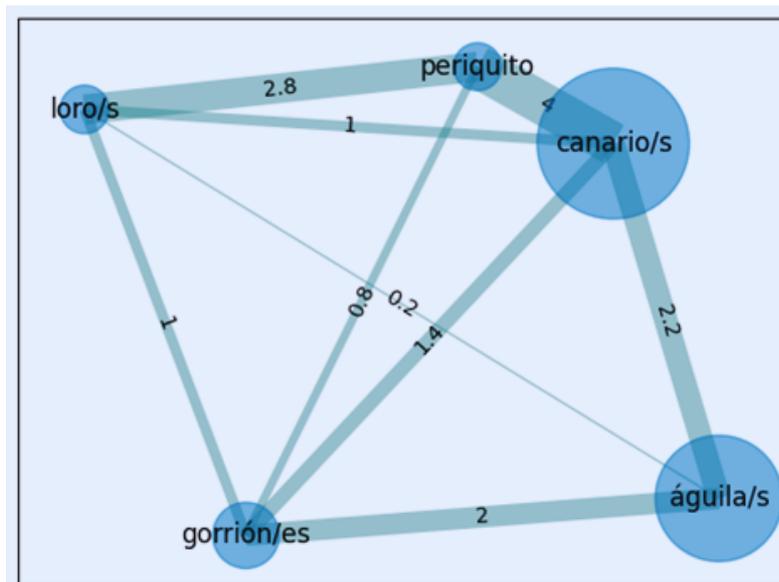
19. Una vez generado un determinado grafo, se puede seleccionar una o varias palabras y obtener su red particular en un SUBGRAFO. En el recuadro inferior derecho llamado TRAZAR SUBGRAFO, aparecen los vocablos/nodos que contiene el gráfico ordenados alfabéticamente. Puede seleccionar uno o más de ellos y crear un grafo solo de la red generada por esas palabras. Determine el NÚMERO DE PASOS PARA INCLUIR VOCABLOS y seleccione CREAR SUBGRAFO.

La imagen siguiente muestra el subgrafo de todas las conexiones directas (1 paso) de la palabra *araña/s* (Hernández Muñoz, 2006 [3]). En un subgrafo de 2 pasos se mostrarían todos los nodos que conectan con *araña/s* a una distancia máxima de dos aristas.



En un subgrafo de 0 pasos solo aparecerán las palabras seleccionadas en la lista. Por ejemplo, en el que aparece a continuación se han seleccionado las aves (*loro*, *periquito*, *canario*, *águila* y *gorrión*) que aparecían en el listado del grafo de las 30 palabras más disponibles de *Animales* (Hernández Muñoz, 2006 [3]).





20. Al generar un grafo se calculan los índices matemáticos de los nodos y la red. En los cuadros de la derecha, DATOS DE LOS NODOS y DATOS DE LA RED, puede encontrar los índices derivados de cada uno de los grafos creados. Si desplaza el cursor sobre ellos, aparece la definición y el cálculo de cada uno de los índices consignados. Son los siguientes:

Datos de los nodos:

Componente conexo
(*Connected component*):

componente al que pertenece el nodo. Los componentes conexos son los subgrafos conectados que se distinguen en un grafo no conectado.

Grado (*Degree*):

número de aristas que conectan con el nodo.

Longitud del camino
(*Path Length*):

promedio de las distancias mínimas que separan el nodo de los otros con los que está conectado.

Grado de entrada
(*In-degree*):

en los grafos dirigidos, número de aristas que llegan al nodo.

Grado de salida
(*Out-degree*):

en los grafos dirigidos, número de aristas que parten del nodo.

Longitud del camino de
entrada:

en los grafos dirigidos, promedio de los caminos más cortos que llegan al nodo.

Longitud del camino de
salida:

en los grafos dirigidos, promedio de los caminos más cortos que parten del nodo.

Datos de los nodos:

Coefficiente de agrupamiento (<i>Clustering Coefficient</i>):	probabilidad de que dos vecinos del nodo sean también vecinos entre sí.
Centralidad de entrada:	en los grafos dirigidos, fracción de los nodos totales con conexiones hacia el nodo.
Centralidad de salida:	en los grafos dirigidos, fracción de los nodos totales con conexiones desde el nodo.
Centralidad:	fracción de los nodos totales a los que está conectado el nodo.

Datos de la red:

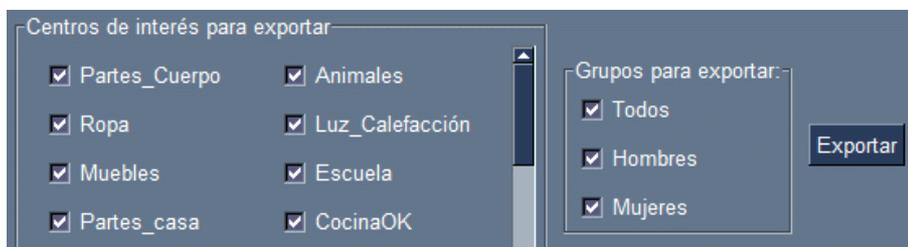
Promedio del grado (<i>Average Degree</i>):	promedio de las conexiones que presenta la red en relación con el número de nodos.
Promedio de la longitud del camino (<i>Average Path Length</i>):	promedio de los caminos más cortos para todos los nodos de la red.
Coefficiente de agrupamiento (<i>Clustering Coefficient</i>):	probabilidad de que dos vecinos de un nodo elegido al azar sean también vecinos entre sí.
Sigma (S):	comparación del agrupamiento y el camino medio de la red con el de una red aleatoria de grado equivalente a través del coeficiente descrito en Humphries y Gurney (2008)[5].
Diámetro:	distancia máxima entre los dos nodos más alejados de la red.
Modularidad:	separación en comunidades de los miembros de la red según se describe en Newman (2011) y Clauset <i>et al.</i> (2004) [6].

[5] Humphries, M. D. y K. Gurney (2008). "Network 'Small-World-Ness': A Quantitative Method for Determining Canonical Network Equivalence", *PLoS ONE*, 3(4).

[6] Newman, M. E. J. (2011). "Networks: An Introduction". Oxford: Oxford University Press.
Clauset, A., M. E. J. Newman y C. Moore (2004). "Finding community structure in very large networks", *Phys. Rev.*, 70.6.

PARA EXPORTAR LOS DATOS

21. Las pestañas GLOBAL, INFORMANTES, DISPONIBILIDAD, COMPARACIÓN y GRAFOS poseen un botón que permite la exportación de los datos que aparecen en pantalla a un fichero Excel. En las pestañas de GLOBAL e INFORMANTES se exportan todos los datos, pero en DISPONIBILIDAD puede seleccionar los listados que desea exportar por centro de interés y por grupo:



22. En la pestaña de GRAFOS, además de exportar a Excel los datos de los nodos y los de la red, tiene la opción de exportar los grafos a Ghephi.

Hernández Muñoz, Natividad, Carmela Tomé Cornejo, Miguel López García y José Antonio Bartol Hernández (2023): *Manual de LexPro*. <https://dispogram.usal.es/>
[Consulta: fecha]



VNiVERSIDAD
D SALAMANCA



UNIVERSITAS
Miguel Hernández

Proyecto PID2020-120436GB-I00 financiado por:



MINISTERIO
DE CIENCIA
E INNOVACIÓN

