

DISPOCEN. MUCHO MÁS QUE UN PROGRAMA PARA EL CÁLCULO DE LA DISPONIBILIDAD LÉXICA

DISPOCEN. MUCH MORE THAN A PROGRAM ABOUT LEXICAL AVAILABILITY

ANTONIO MANUEL ÁVILA-MUÑOZ

Universidad de Málaga, España

amavila@uma.es

<https://orcid.org/0000-0002-5239-2670>

JOSÉ MARÍA SÁNCHEZ SÁEZ

Universidad de Málaga, España

jmss@uma.es

<https://orcid.org/0000-0003-4707-5029>

NANA ODISHELIDZE

Iv. Javakhishvili Tbilisi State University, Georgia

nana.odishelidze@tsu.ge

<https://orcid.org/0000-0001-8186-7105>

Resumen

DispoCen es un sistema para el análisis de la disponibilidad y la centralidad léxica. Aunque existen programas específicos para el cálculo de los citados índices, estos suelen restringir en exceso las posibilidades de análisis y explotación de los datos, bien porque se trata de herramientas obsoletas, bien porque sus códigos son excesivamente cerrados e inaccesibles. *DispoCen* está basado en una librería de herramientas en

Abstract

DispoCen is a system for the analysis of availability and lexical centrality. Although there are specific programs for calculating the mentioned index, these tend to excessively restrict the possibilities of data analysis, either because they are obsolete tools, or because their codes are excessively closed and inaccessible. *DispoCen* is based on a library of tools in *R* that makes the development of multiple applications and

Este trabajo ha sido posible gracias a la financiación y patrocinio del Ministerio de Ciencia, Innovación y Universidades al Proyecto de Investigación Agenda 2050. El español de Málaga: procesos de variación y cambio espaciales y sociales (PID2019-104982GB-C5-2).

Para citar este artículo: Ávila-Muñoz, Antonio Manuel; Sánchez Sáez, José María y Odishelidze, Nana (2021). *Dispocen*. Mucho más que un programa para el cálculo de la disponibilidad léxica. *ELUA*, 35: 9-36. <https://doi.org/10.14198/ELUA2021.35.1>

Recibido: 15/10/2020, Aceptado: 16/12/2020

© 2021 Antonio Manuel Ávila-Muñoz, José María Sánchez Sáez, Nana Odishelidze



Este trabajo está sujeto a una licencia de **Reconocimiento 4.0 Internacional de Creative Commons (CC BY 4.0)**

R que pone al alcance de quienes estudian el léxico el desarrollo de múltiples aplicaciones y modelos originales. En este trabajo hemos incluido los códigos necesarios para ejecutar los análisis, con lo que potenciamos la necesaria replicabilidad que favorece el trabajo autónomo de la comunidad investigadora. Para facilitar el acceso al sistema, también presentamos una sencilla utilidad gráfica que permite el acceso a los análisis más usuales. Como muestra de las posibilidades de *DispoCen*, incluimos un apartado específico con propuestas de análisis realizadas con filtros sociológicos.

PALABRAS CLAVE: disponibilidad léxica; centralidad léxica; *R*; herramientas de análisis; *DispoCen*.

original models to those who study the lexicon. In this paper we have included the necessary codes to run the analysis, thereby enhancing the necessary replicability that allows the autonomous work of the research community. To facilitate access to the system, we also present a simple graphical tool that facilitates access to the most common analyzes. As a sample of the possibilities of *DispoCen*, we include a specific section with proposals for analysis made with sociological items.

KEYWORDS: lexical availability; lexical centrality; *R*; analysis tools; *DispoCen*.

1. INTRODUCCIÓN

Este documento es una guía rápida de uso del programa *DispoCen* para el cálculo de la disponibilidad léxica, la centralidad léxica y las diferentes funciones que se derivan de ambas. Hemos creado una librería de herramientas en *R* que permite, a través de programación funcional, implementar múltiples modelos y aplicaciones, entre las que destaca, precisamente, el desarrollo inmediato de las funcionalidades señaladas (disponibilidad, centralidad). Sin embargo, tal y como se demuestra en este artículo, es preciso hacer notar que estas funciones de aplicación directa representan tan solo un subconjunto de las posibilidades que el sistema ofrece.

Hemos obviado conscientemente la exposición de los fundamentos teóricos y metodológicos de la disponibilidad y la centralidad léxica, pues este documento va dirigido a investigadores especializados en esta materia. Consideramos que la comunidad investigadora demandaba un contexto actualizado a partir del cual realizar sus análisis y organizar los datos de manera autónoma. La obsolescencia de algunos programas previos para el cálculo de la disponibilidad, o el acceso restringido a otros, justifica la necesidad de una herramienta como la que presentamos en este trabajo.

Una de las ventajas del sistema que se ha creado es que nos ha permitido incluir en el mismo documento el análisis realizado junto con el código necesario para su ejecución, con lo que se potencia la replicabilidad, característica cada vez más exigida en el ámbito académico. Se consigue así que otros investigadores puedan replicar, tanto para verificar como para desarrollar, el trabajo expuesto. En nuestro caso particular, este objetivo se cumple al exponer en el mismo documento la disertación y el código con el que se han realizado los análisis.

Tras la exposición de la naturaleza del sistema creado, el presente artículo orienta al lector en el proceso de instalación y ejecución del entorno propuesto mediante ejemplos y demostraciones que justifican cada uno de los pasos realizados. Para facilitar el proceso de instalación y ejecución de *DispoCen*, hemos alojado un vídeo explicativo en *Youtube* donde se detallan cada uno de los pasos necesarios para que la herramienta pueda usarse de manera eficaz (<https://m.youtube.com/watch?v=IU5VfUvG4Ag>).

2. EXPOSICIÓN, JUSTIFICACIÓN Y CONTEXTUALIZACIÓN DEL SISTEMA

El sistema se ha desarrollado en *R* (<https://www.r-project.org/>), una implementación abierta del lenguaje de programación *S*, que lleva en desarrollo desde 1992 con una primera versión estable desde el año 2000. Es fácilmente extensible y permite la reutilización de códigos y datos. Se trata de un lenguaje soportado por una amplia comunidad a lo largo del mundo que se emplea de forma habitual en entornos de análisis de datos de ingeniería y medicina.

Rstudio (<https://rstudio.com/>) es un entorno comercial de desarrollo para *R* que, no obstante, permite el uso personal gratuito y fomenta la productividad en el uso de este lenguaje. El sistema está disponible para la gran mayoría de plataformas y sistemas operativos. En nuestro caso, para facilitar el trabajo, se proporcionan utilidades que pueden ejecutarse dentro de este entorno y que facilitan el proceso de análisis. Es posible utilizar las funcionalidades proporcionadas por nuestro software sin el empleo de *RStudio*, aunque consideramos que su uso se haría innecesariamente complejo y se perderían muchas de las posibilidades que proporciona este marco de trabajo. Dentro del sistema construido en *RStudio*, nos resulta especialmente interesante, para los objetivos didácticos de este trabajo, la posibilidad de desarrollar el paradigma de la ‘programación literaria’, según el cual, en un mismo documento se van integrando los contenidos textuales —como la exposición de hechos o la interpretación de los análisis— con bloques de código que llevan a cabo las manipulaciones y representaciones de datos.

A diferencia de soluciones basadas en programas autónomos, la implementación de esta alternativa dentro de un sistema como el formado por *R* junto con *RStudio* permite, además, la integración de todo el sistema de gestión de datos. De esta manera, los resultados obtenidos a partir de las herramientas proporcionadas se incorporan al sistema, con lo que se proporciona la integración de datos en nuevos procesos y la aplicación personal de los resultados que se obtienen en función de las demandas específicas de los usuarios. En este documento se pueden encontrar algunos de estos desarrollos particulares que pueden servir de ejemplo.

La distribución de las soluciones es otro problema que ya ha encontrado respuesta eficiente dentro de la comunidad de usuarios de *R*. Repositorios de código como *GitHub* permiten la gestión y el control automatizado de las diferentes versiones. Es posible llevar a cabo el control exhaustivo de las modificaciones que se van realizando, sin necesidad de descargar e instalar continuamente diferentes actualizaciones de los mismos programas. Con dos líneas de código se asegura la instalación de la última versión disponible. Además, este modelo de intercambio de código permite que los proyectos no desaparezcan, ya que se pueden generar ramas de cualquier repositorio de código para continuar su desarrollo, independientemente de los creadores originales.

Entre las múltiples herramientas construidas sobre *R* hemos encontrado muy interesantes y productivas las que conforman el *Universo Tidyverse* (<https://www.tidyverse.org/>). Se trata de un conjunto de librerías que facilitan y permiten una manipulación expresiva de los datos, basada en el concepto de ‘tubería’. Este tratamiento se construye mediante la secuenciación de manipulaciones que van encadenándose. De esta manera, el resultado de una etapa de análisis constituye la fuente de datos para la siguiente. Esta característica, junto con operadores adaptados a este tipo de trabajo, proporciona una forma potente y cómoda de llevar a cabo transformaciones que, de otra forma, serían muy complejas. El tratamiento de datos presentado en este trabajo se realiza utilizando estas herramientas.

3. INSTALACIÓN DE *DISPOCEN*

3.1. Requisitos previos

En primer lugar, es necesario instalar el sistema *R*. Para ello, se accede desde la sección *Download* de la página del proyecto, <https://www.r-project.org/>, al enlace CRAN. Desde aquí se nos permite seleccionar la *Red Iris*, que es una réplica local de *R* específica para España (<https://cran.rediris.es/>). Desde la página principal seleccionamos el enlace adecuado para descargar la versión correspondiente a nuestro sistema operativo. En el caso de elegir la versión para Windows, se debe seleccionar la opción denominada ‘base’. Finalmente, accedemos al binario que instala el sistema que, en el momento de redacción de estas líneas (febrero de 2021), corresponde a la versión 4.0.4, en el enlace (<https://cran.rediris.es/bin/windows/base/R-4.0.4-win.exe>)[<https://cran.rediris.es/bin/windows/base/R-4.0.4-win.exe>]. Para poder instalar ciertas actualizaciones y paquetes es conveniente instalar también el paquete *Rtools*, que proporciona a *R* la capacidad de compilar paquetes nativos. Se puede encontrar en <https://cran.r-project.org/bin/windows/Rtools/>, y la versión disponible en este momento es https://cran.r-project.org/bin/windows/Rtools/rtools40-x86_64.exe.

A continuación, se recomienda aunque no es estrictamente necesario instalar el entorno de productividad *RStudio* (<https://rstudio.com/>) [<https://rstudio.com/>]. A través de la ruta *Product > Open Source > RStudio*, se llega a la página de descripción del producto, del que existen dos versiones: *Desktop* y *Server*. En nuestro caso, usaremos *Desktop*, en su modalidad *Open Source Edition*, que se puede descargar en (<https://rstudio.com/products/rstudio/download/>)[<https://rstudio.com/products/rstudio/download/>]. La versión actual para Windows se encuentra en (<https://download1.rstudio.org/desktop/windows/RStudio-1.3.1073.exe>)[<https://download1.rstudio.org/desktop/windows/RStudio-1.3.1073.exe>].

3.2. Instalación

Una vez instalada la aplicación *RStudio*, para poder utilizar las herramientas que proponemos, el usuario debe llevar a cabo un segundo paso de instalación que, en nuestra opinión, no es más complejo que cualquier proceso de instalación en el sistema *R*. Este paso ha de realizarse únicamente cuando se pretenda instalar o actualizar el paquete en el sistema. Puesto que nuestra intención es seguir trabajando en la herramienta, sería recomendable realizar esta acción de forma periódica.

Desde la consola de *R*, las siguientes órdenes instalan el paquete *DispoCen* desde el repositorio *GitHub*. Todo el código está implementado en *R*, con lo que es fácil de revisar.

```
install.packages("devtools")
library(devtools)
install_github("jmss70/disprocen")
```

Para facilitar la introducción de la herramienta, se han impuesto como requerimientos tanto las utilidades del universo *Tidyverse* —que facilitan el procesamiento de datos y se usa de forma extensiva en este y otros trabajos— como la librería *flextable* y *kableExtra*, con el objetivo de mejorar la presentación de las tablas.

Una vez que se han llevado a cabo los pasos anteriores, y si no se ha producido ningún contratiempo, el sistema está preparado para ejecutar el cálculo de los índices de disponibilidad y centralidad junto con sus funcionalidades.

3.2.1. Carga de librerías y datos

Para poder usar las librerías es preciso cargarlas en nuestra sesión de trabajo. De esta manera se ponen a disposición del usuario las funciones que proporcionan.

library(dispocen)

El formato requerido de los datos coincide con el tradicionalmente usado por los programas más empleados hasta ahora para el cálculo de la disponibilidad léxica (*Lexidisp*, *Dispolex*, esencialmente). Sin embargo, consideramos que parte de la información que contiene este formato es redundante y sería conveniente estudiar, en un futuro cercano, la posibilidad de establecer un estándar de codificación que sea más coherente con los modelos de datos normalizados.

De momento, se espera que los datos estén en un archivo de texto, con campos separados por espacios, según la estructura propuesta en el diseño del programa *Lexidisp* (Moreno Fernández et al. 1995):

1. Un campo con información sociológica básica del informante formado por cinco caracteres sucesivos. Cada uno de ellos representa la codificación de una variante.
2. Un campo de identificación del usuario de tres caracteres.
3. Un campo de identificación del centro de interés de dos caracteres.
4. Una lista de palabras separadas por comas que sigue el orden de su aparición en las listas originales.

Un ejemplo de dos líneas sería:

```
21131 001 01 mano, pie, brazo, cerebro, pulmón, nariz, extremidad, ojo, boca, diente, pelo, oreja,
culo, vagina
12131 002 01 riñón, corazón, garganta, cabeza, pierna, pie, hígado, estómago, mano, brazo, ante-
brazo, abdomen, pecho, ojo, boca, oído, dedo, rodilla, costilla.
```

Tanto los datos que hemos presentado en el ejemplo anterior, como el resto de datos que usamos para la presentación de *DispoCen* en el presente trabajo proceden del corpus textual recogido y expuesto en el libro de Ávila Muñoz y Villena Ponsoda (eds.) (2010) utilizado para analizar los condicionantes sociales de la ciudad de Málaga (N=72 informantes malagueños).

Suponiendo que tenemos todos los datos cargados en un archivo, denominado *datos.txt*, que estará alojado en el mismo directorio que el *script* de procesamiento, se podrían cargar los datos como:

```
data <- read.dispocen("datos.txt")
data %>%
  head() %>%
  flextable() %>% autofit() %>%
```

theme_booktabs()

| Información sociológica | Sujetos | Centros | Palabras |
|-------------------------|---------|---------|--------------|
| 21131 | 001 | 01 | mano, pi.... |
| 12131 | 002 | 01 | riñón, c.... |
| 12213 | 003 | 01 | brazo, m.... |
| 22214 | 004 | 01 | brazo, o.... |
| 12214 | 005 | 01 | cabeza, |
| 22213 | 006 | 01 | pie, man.... |

3.3. Cálculo de la disponibilidad

La función general del cálculo de la disponibilidad es *build.availability*. Sin embargo, el empleo de esta función requiere del uso de varios parámetros. Se han construido dos funciones de utilidad que encapsulan su ejecución y ofrecen los dos modelos que consideramos, por el momento, más interesantes (modelo López-Strassburger y modelo Ávila-Sánchez). Aunque Callealta Barroso y Gallego Gallego (2016) propusieron nuevas fórmulas para el cálculo de esta función a partir de una revisión crítica de los modelos tradicionales, en este trabajo no son consideradas al tratarse de propuestas excesivamente controladas y poco realistas que han tenido menos recorrido que las de López-Strassburger y Ávila-Sánchez. A continuación, mostramos su empleo mediante ejemplos.

3.3.1. Modelo de López-Strassburger

Este modelo, que es el más conocido y el último de una serie de iteraciones para la cuantificación de la disponibilidad, se puede usar como:

```
disponibilidad <- build.lopezstrass.availability(data)
```

El resultado es un nuevo marco de datos cuyos campos son el centro de interés, la palabra y la disponibilidad calculada, junto con otros datos necesarios para su obtención como la posición y las frecuencias relativas y acumuladas dentro del centro de interés:

```
head(disponibilidad) %>%
  flextable() %>%
  set_header_labels(centers = "Centro de interés",
                    word s= "Palabra",
                    order = "Orden",
                    availability = "Disponibilidad",
                    freq.abs = "Frecuencia absoluta",
                    freq.abs.cum = "Frecuencia absoluta acumulada",
                    freq.rel = "Frecuencia relativa",
                    freq.rel.cum = "Frecuencia relativa acumulada") %>%
  colformat_double(j=c(4,6,8), digits=6) %>%
  width(j=c(1,3,5,7),width=.65) %>%
```

```
width(j=c(2,4,6,8),width=.9) %>%
theme_booktabs()
```

| Centro | Palabra | Orden | Disponibilidad | Frec. absoluta | Frec. relativa | Frec. absoluta acumulada | Frec. relativa acumulada |
|--------|---------|-------|----------------|----------------|----------------|--------------------------|--------------------------|
| 01 | ojo | 1 | 0.632987 | 58 | 0.037395 | 58 | 0.037395 |
| 01 | brazo | 2 | 0.551808 | 50 | 0.032237 | 108 | 0.069632 |
| 01 | pierna | 3 | 0.507215 | 49 | 0.031593 | 157 | 0.101225 |
| 01 | mano | 4 | 0.489482 | 42 | 0.027079 | 199 | 0.128304 |
| 01 | cabeza | 5 | 0.478054 | 37 | 0.023856 | 236 | 0.152160 |
| 01 | pie | 6 | 0.450128 | 43 | 0.027724 | 279 | 0.179884 |

Tabla 1. Resultado del análisis según el modelo López-Strassburger. Centro de interés 1. Fuente: Ávila y Villena (eds.) (2010)

Al ser un marco de datos de *R* estándar, se pueden realizar sobre él todas las operaciones que permite el sistema. En el siguiente caso, por ejemplo, hemos seleccionado los resultados obtenidos en el centro de interés 03. *Partes de la casa*, los hemos organizado en orden decreciente de disponibilidad y mostramos los 10 términos más disponibles en la Tabla 2:

```
disponibilidad %>%
  filter(centers=="03") %>%
  arrange(-availability) %>%
  head(10) %>%
  flextable() %>%
  set_header_labels(centers = "Centro de interés",
                    words="Palabra",
                    order="Orden",
                    availability="Disponibilidad",
                    freq.abs="Frecuencia absoluta",
                    freq.abs.cum="Frecuencia absoluta acumulada",
                    freq.rel="Frecuencia relativa",
                    freq.rel.cum="Frecuencia relativa acumulada") %>%
  colformat_double(j=c(4,6,8), digits=6) %>%
  width(j=c(1,3,5,7),width=.65) %>%
  width(j=c(2,4,6,8),width=.9) %>%
  theme_booktabs()
```

| Centro | Palabra | Orden | Disponibilidad | Frec. absoluta | Frec. relativa | Frec. absoluta acumulada | Frec. relativa acumulada |
|--------|------------------|-------|----------------|----------------|----------------|--------------------------|--------------------------|
| 03 | cocina | 1 | 0.841569 | 69 | 0.066925 | 69 | 0.066925 |
| 03 | (cuarto de) baño | 2 | 0.746670 | 66 | 0.064016 | 135 | 0.130941 |
| 03 | salón | 3 | 0.715562 | 58 | 0.056256 | 193 | 0.187197 |
| 03 | dormitorio | 4 | 0.636728 | 55 | 0.053346 | 248 | 0.240543 |

| Centro | Palabra | Orden | Disponibilidad | Frec. absoluta | Frec. relativa | Frec. absoluta acumulada | Frec. relativa acumulada |
|--------|------------------|-------|----------------|----------------|----------------|--------------------------|--------------------------|
| 03 | terrace | 5 | 0.608507 | 62 | 0.060136 | 310 | 0.300679 |
| 03 | comedor | 6 | 0.439152 | 39 | 0.037827 | 349 | 0.338506 |
| 03 | salita | 7 | 0.337618 | 31 | 0.030068 | 380 | 0.368574 |
| 03 | lavadero | 8 | 0.318056 | 33 | 0.032008 | 413 | 0.400582 |
| 03 | pasillo | 9 | 0.310978 | 34 | 0.032978 | 447 | 0.433560 |
| 03 | (cuarto de aseo) | 10 | 0.258557 | 24 | 0.023278 | 471 | 0.456838 |

Tabla 2. Resultado del análisis según el modelo López-Strassburger según criterios de filtrado. Fuente: Ávila y Villena (eds.) (2010)

También es posible utilizar los mecanismos de *R* para representar la curva de disponibilidad, esto es, la sucesión de valores de disponibilidad, una vez ordenados en valor decreciente de disponibilidad. El Gráfico 1 muestra la curva resultante para el centro de interés 3:

```

disponibilidad %>%
  filter(centers == "03") %>%
  arrange(-availability) %>%
  ggplot(aes(x=order, y=availability)) + geom_line() +
  xlab("Sucesión de palabras") + ylab("Disponibilidad")

```

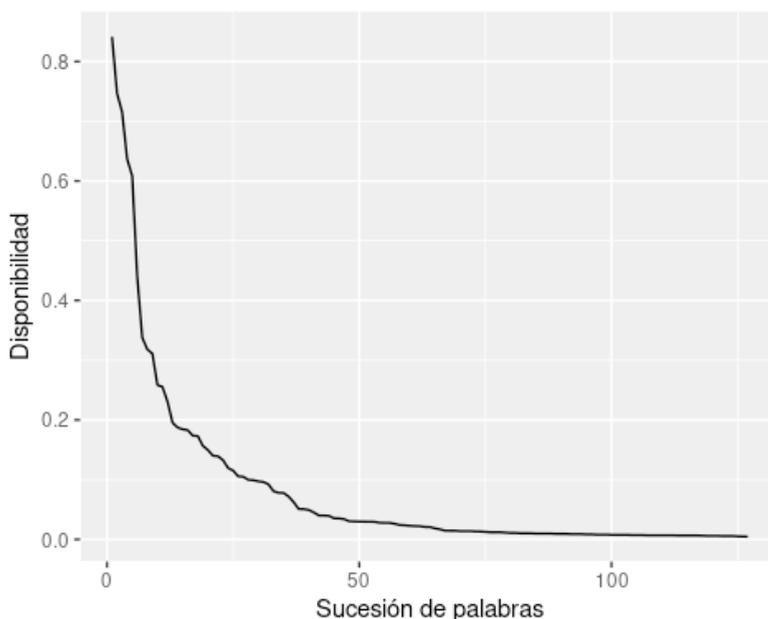


Gráfico 1. Curva de disponibilidad según el modelo López-Strassburger. Centro de interés 3. Fuente: Ávila y Villena (eds.) (2010)

3.3.2. Modelo de Ávila-Sánchez

Ávila y Sánchez (2014) propusieron un macro-modelo para el estudio de la disponibilidad a partir de la Teoría de los Conjuntos Difusos y mediante la modelización de los conceptos que se pretenden representar. Esta modelización se produce en dos etapas: en la primera, se cuantifica la relevancia de cada término en las pruebas obtenidas para cada hablante y centro de interés según una ley descendente a medida que se avanza en cada listado; en la segunda etapa, se integra esa información con una ley aditiva que considera los distintos valores alcanzados para cada palabra en cada centro de interés.

Hay múltiples elecciones posibles, pero en distintas pruebas parecieron más productivas las que utilizaban en la primera etapa una ley de Zipf-Mandelbrot y en la segunda una adición probabilística. La interpretación de los valores obtenidos corresponde a la ‘centralidad’ de cada término en cada centro de interés. Un valor de 1, o muy cercano, respondería a la pertenencia al núcleo del vocabulario específico del centro de interés, mientras que un valor próximo a 0 indicaría que se trataría de un término poco accesible y, por tanto, des-centralizado.

La función que lleva a cabo este análisis, *build.avilasanchez.availability*, se utiliza de la misma forma que la mostrada anteriormente en el modelo López-Strassburger:

```
disponibilidad <- build.avilasanchez.availability(data)
```

El resultado es, de igual modo, un nuevo marco de datos con la disponibilidad de cada término en cada centro de interés

```
disponibilidad %>%
  filter(centers=="03") %>%
  arrange(-availability) %>%
  head(10) %>%
  flextable() %>%
  set_header_labels(centers = "Centro de interés",
                    words = "Palabra",
                    order = "Orden",
                    availability = "Disponibilidad",
                    freq.abs = "Frecuencia absoluta",
                    freq.abs.cum = "Frecuencia absoluta acumulada",
                    freq.rel = "Frecuencia relativa",
                    freq.rel.cum = "Frecuencia relativa acumulada") %>%
  colformat_double(j=c(4,6,8), digits=6) %>%
  width(j=c(1,3,5,7),width=.65) %>%
  width(j=c(2,4,6,8),width=.9) %>%
  theme_booktabs()
```

| Centro | Palabra | Orden | Disponibilidad | Frec. absoluta | Frec. relativa | Frec. absoluta acumulada | Frec. relativa acumulada |
|--------|------------------|-------|----------------|----------------|----------------|--------------------------|--------------------------|
| 03 | cocina | 1 | 1.000000 | 69 | 0.066925 | 69 | 0.066925 |
| 03 | salón | 2 | 1.000000 | 58 | 0.056256 | 127 | 0.123181 |
| 03 | (cuarto de) baño | 3 | 1.000000 | 66 | 0.064016 | 193 | 0.187197 |
| 03 | dormitorio | 4 | 1.000000 | 55 | 0.053346 | 248 | 0.240543 |
| 03 | comedor | 5 | 0.999928 | 39 | 0.037827 | 287 | 0.278371 |
| 03 | terracea | 6 | 0.999900 | 62 | 0.060136 | 349 | 0.338506 |
| 03 | salita | 7 | 0.997438 | 31 | 0.030068 | 380 | 0.368574 |
| 03 | entrada | 8 | 0.996197 | 23 | 0.022308 | 403 | 0.390883 |
| 03 | (cuarto de) aseo | 9 | 0.994715 | 24 | 0.023278 | 427 | 0.414161 |
| 03 | pasillo | 10 | 0.988474 | 34 | 0.032978 | 461 | 0.447139 |

Tabla 3. Resultado del análisis según el modelo Ávila-Sánchez. Centro de interés 3. Fuente: Ávila y Villena (eds.) (2010)

Como es lógico, el resultado se puede procesar como cualquier marco de datos del sistema *R*:

```

disponibilidad %>%
  filter(centers=="03") %>%
  arrange(-availability) %>%
  ggplot(aes(x=order, y=availability)) + geom_line() +
  xlab("Sucesión de palabras") + ylab("Disponibilidad")

```

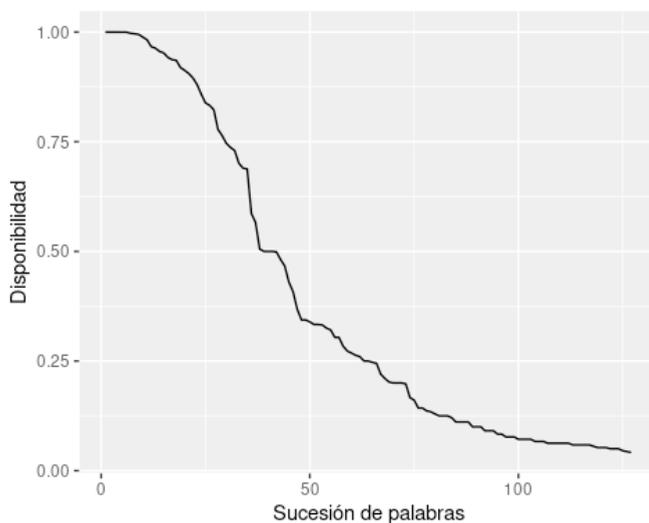


Gráfico 2. Curva de disponibilidad según el modelo Ávila-Sánchez. Centro de interés 3. Fuente: Ávila y Villena (eds.) (2010)

Debido a las características de los operadores aditivos de la teoría de los conjuntos difusos es posible, como en el ejemplo anterior, que la forma de la curva no sea idéntica a la obtenida según el modelo de López-Strassburger: esto es debido a que quizás no aparezcan términos con valores cercanos a 1 (por ejemplo, si se tienen pocos datos y estos son relativamente dispersos) o, por el contrario, demasiados términos con valoraciones cercanas a 1 (por ejemplo, si se trabaja con muchas muestras). Aunque las opciones anteriores son improbables, pero posibles, hemos optado por regular la curva mediante un parámetro adicional (k) que modifica su trayectoria (ascendente-descendente), pero manteniendo su forma y clasificación. El valor por defecto de k es 1. Si se da un valor entre 0 y 1, la curva bajará (menos términos con valores cercanos a 1), mientras que si a k se le da un valor mayor que la unidad la curva subirá (más términos cercanos a la unidad).

```
disponibilidad <- build.avilasanchez.availability(data, k = 0.1)
```

```
disponibilidad %>%
  filter(centers == "03") %>%
  arrange(-availability) %>%
  ggplot(aes(x=order, y=availability)) + geom_line() +
  xlab("Sucesión de palabras") + ylab("Disponibilidad")
```

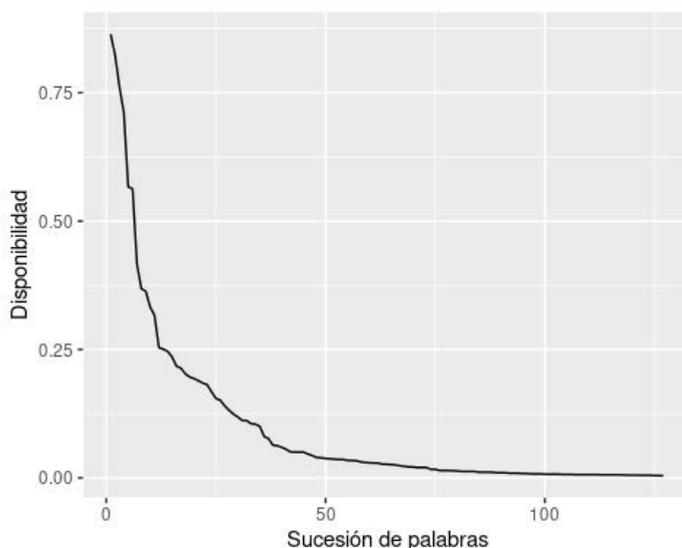


Gráfico 3. Curva de disponibilidad subestimada según el modelo Ávila-Sánchez. Centro de interés 3. Fuente: Ávila y Villena (eds.) (2010)

El resultado final será, en cualquier caso, un nuevo marco de datos con la disponibilidad de cada término en cada centro de interés

```
disponibilidad <- build.avilasanchez.availability(data, k = 2)
```

```

disponibilidad %>%
  filter(centers == "03") %>%
  arrange(-availability) %>%
  ggplot(aes(x=seq_along(availability), y=availability)) + geom_line() +
  xlab("Sucesión de palabras") + ylab("Disponibilidad")

```

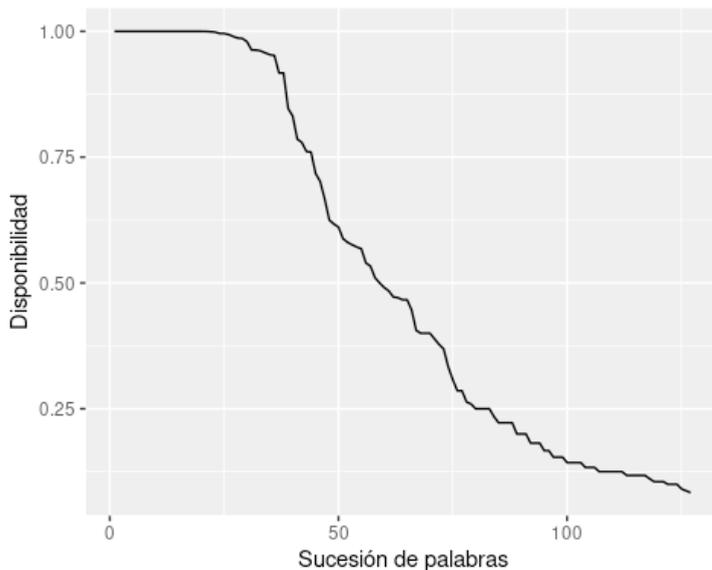


Gráfico 4. Curva de disponibilidad sobreestimada según el modelo Ávila-Sánchez. Centro de interés 3. Fuente: Ávila y Villena (eds.) (2010)

Téngase en cuenta que los ejemplos mostrados representan de forma artificial, manipulada y consciente, valores muy extremos usados a modo de exposición. En nuestras pruebas, realizadas siempre con datos reales procedentes de investigaciones previas, hemos encontrado que el valor de referencia es lo suficientemente fiable en todos los casos que hemos analizado.

3.4. Niveles de disponibilidad

Una pregunta recurrente en algunos estudios previos de disponibilidad léxica es considerar cuál es el tamaño del conjunto de elementos que se debería establecer para configurar el núcleo de un centro de interés (Samper 1999: 554-555; Bartol 2001: 227-230; Carcedo 2001: 62). Para responder a esta cuestión, y a partir del marco teórico creado en nuestra propuesta, se proporciona una herramienta que etiqueta los términos por niveles de centralidad. El nivel 0 correspondería a aquellos elementos que no pertenecen al núcleo, es decir, aquellos términos que no son generalmente accesibles para el conjunto analizado. Los niveles 1, 2, 3 y sucesivos representarían un mayor grado de centralidad y aproximación al centro de interés. La Tabla 4 muestra los elementos más nucleares en cada centro de interés.

```

disponibilidad <- build.avilasanchez.availability(data)
levels <- classify.availability.levels(disponibilidad)
levels %>%
  head(20) %>%
  arrange(-availability) %>%
  select(-order) %>%
  flextable() %>%
  set_header_labels(centers = "Centro de interés",
                    words = "Palabra",
                    availability = "Disponibilidad",
                    level = "Nivel de disponibilidad",
                    cutlevel = "Nivel de corte",
                    freq.abs = "Frecuencia absoluta",
                    freq.abs.cum = "Frecuencia absoluta acumulada",
                    freq.rel = "Frecuencia relativa",
                    freq.rel.cum = "Frecuencia relativa acumulada") %>%
  colformat_double(j=c(5,7,9), digits=5) %>%
  width(j=c(1,4,6,8),width=.65) %>%
  width(j=c(2,3,5,6,7,9),width=.75) %>%
  theme_booktabs()

```

| Centro | Palabra | Disponibilidad | Nivel de disponibilidad | Nivel de corte | Frec. absoluta | Frec. relativa | Frec. absoluta acumulada | Frec. relativa acumulada |
|--------|-------------------------|----------------|-------------------------|----------------|----------------|----------------|--------------------------|--------------------------|
| 14 | perro | 1 | 6 | 0.94971 | 67 | 0.04206 | 67 | 0.04206 |
| 12 | coche | 1 | 6 | 0.95891 | 65 | 0.05717 | 65 | 0.05717 |
| 19 | rojo | 1 | 6 | 0.96142 | 66 | 0.05255 | 66 | 0.05255 |
| 03 | cocina | 1 | 6 | 0.96381 | 69 | 0.06693 | 69 | 0.06693 |
| 19 | azul | 1 | 6 | 0.96142 | 68 | 0.05414 | 134 | 0.10908 |
| 14 | gato | 1 | 6 | 0.94971 | 64 | 0.04018 | 131 | 0.08223 |
| 04 | mesa | 1 | 6 | 0.95622 | 54 | 0.05182 | 54 | 0.05182 |
| 03 | salón | 1 | 6 | 0.96381 | 58 | 0.05626 | 127 | 0.12318 |
| 17 | euro | 1 | 6 | 0.94392 | 45 | 0.04014 | 45 | 0.04014 |
| 09 | aire acondi- cionado | 1 | 6 | 0.95832 | 58 | 0.07827 | 58 | 0.07827 |
| 06 | tenedor | 1 | 6 | 0.95511 | 59 | 0.05530 | 59 | 0.05530 |
| 11 | árbol | 1 | 6 | 0.93846 | 51 | 0.04048 | 51 | 0.04048 |
| 08 | pizarra | 1 | 6 | 0.95052 | 62 | 0.04769 | 62 | 0.04769 |
| 10 | calle | 1 | 6 | 0.95278 | 51 | 0.03783 | 51 | 0.03783 |
| 20 | ola | 1 | 6 | 0.94369 | 46 | 0.03430 | 46 | 0.03430 |
| 04 | silla | 1 | 6 | 0.95622 | 63 | 0.06046 | 117 | 0.11228 |
| 09 | ventilador | 1 | 6 | 0.95832 | 60 | 0.08097 | 118 | 0.15924 |
| 02 | pantalón | 1 | 6 | 0.95766 | 62 | 0.04829 | 62 | 0.04829 |
| 08 | mesa | 1 | 6 | 0.95052 | 46 | 0.03538 | 108 | 0.08308 |
| 01 | cabeza | 1 | 6 | 0.95760 | 37 | 0.02386 | 37 | 0.02386 |

Tabla 4. Elementos más nucleares en cada centro de interés. Fuente: Ávila y Villena (eds.) (2010)

```

levels %>%
  filter(centers=="01") %>%
  arrange(-availability) %>%
  filter(level > 0) %>%
  select(-order) %>%
  flextable() %>%
  set_header_labels(centers = "Centro de interés",
                    words = "Palabra",
                    availability = "Disponibilidad",
                    level = "Nivel de disponibilidad",
                    cutlevel = "Nivel de corte",
                    freq.abs = "Frecuencia absoluta",
                    freq.abs.cum = "Frecuencia absoluta acumulada",
                    freq.rel = "Frecuencia relativa",
                    freq.rel.cum = "Frecuencia relativa acumulada") %>%
  colformat_double(j=c(3,5,7,9), digits=5) %>%
  width(j=c(1,4,6,8), width=.65) %>%
  width(j=c(2,3,5,6,7,9), width=.75) %>%
  theme_booktabs()

```

La Tabla 5 representa en sentido decreciente, y por motivos de espacio, solamente los dos primeros elementos seleccionados por el programa en cada conjunto de corte dentro del centro de interés *1. El cuerpo humano*. Con esta información se puede construir una representación en la que se observa la distribución de las disponibilidades en el centro de interés y los diferentes conjuntos de corte.

| Centro de interés | Palabra | Disponibilidad | Nivel de disponibilidad | Nivel de corte | Frecuencia absoluta | Frecuencia relativa | Frecuencia absoluta acumulada | Frecuencia relativa acumulada |
|-------------------|------------|----------------|-------------------------|----------------|---------------------|---------------------|-------------------------------|-------------------------------|
| 01 | cabeza | 1.00000 | 6 | 0.95760 | 37 | 0.02386 | 37 | 0.02386 |
| 01 | brazo | 1.00000 | 6 | 0.95760 | 50 | 0.03224 | 87 | 0.05609 |
| 01 | ... | ... | 6 | ... | ... | ... | ... | ... |
| 01 | pulmón | 0.92715 | 5 | 0.91971 | 31 | 0.01999 | 645 | 0.41586 |
| 01 | riñón | 0.91971 | 5 | 0.91971 | 23 | 0.01483 | 668 | 0.43069 |
| 01 | ... | ... | 5 | ... | ... | ... | ... | ... |
| 01 | estómago | 0.91195 | 4 | 0.85694 | 25 | 0.01612 | 693 | 0.44681 |
| 01 | extremidad | 0.88274 | 4 | 0.85694 | 8 | 0.00516 | 701 | 0.45197 |
| 01 | ... | ... | 4 | ... | ... | ... | ... | ... |
| 01 | cara | 0.84609 | 3 | 0.76528 | 10 | 0.00645 | 749 | 0.48291 |
| 01 | cerebro | 0.84392 | 3 | 0.76528 | 12 | 0.00774 | 761 | 0.49065 |
| 01 | ... | ... | 3 | ... | ... | ... | ... | ... |
| 01 | cuello | 0.73645 | 2 | 0.61961 | 16 | 0.01032 | 883 | 0.56931 |
| 01 | esqueleto | 0.70370 | 2 | 0.61961 | 4 | 0.00258 | 887 | 0.57189 |

| Centro de interés | Palabra | Disponibilidad | Nivel de disponibilidad | Nivel de corte | Frecuencia absoluta | Frecuencia relativa | Frecuencia absoluta acumulada | Frecuencia relativa acumulada |
|-------------------|------------|----------------|-------------------------|----------------|---------------------|---------------------|-------------------------------|-------------------------------|
| 01 | ... | ... | 2 | ... | ... | ... | ... | ... |
| 01 | vena | 0.61347 | 1 | 0.46263 | 11 | 0.00709 | 974 | 0.62798 |
| 01 | enfermedad | 0.60225 | 1 | 0.46263 | 5 | 0.00322 | 979 | 0.63121 |
| 01 | ... | ... | 1 | ... | ... | ... | ... | ... |

Tabla 5. Muestra de los elementos seleccionados por niveles de corte. Centro de interés 1.
Fuente: Ávila y Villena (eds.) (2010)

levels %>%

```
filter(centers=="04") %>%
```

```
mutate(level=factor(level)) %>%
```

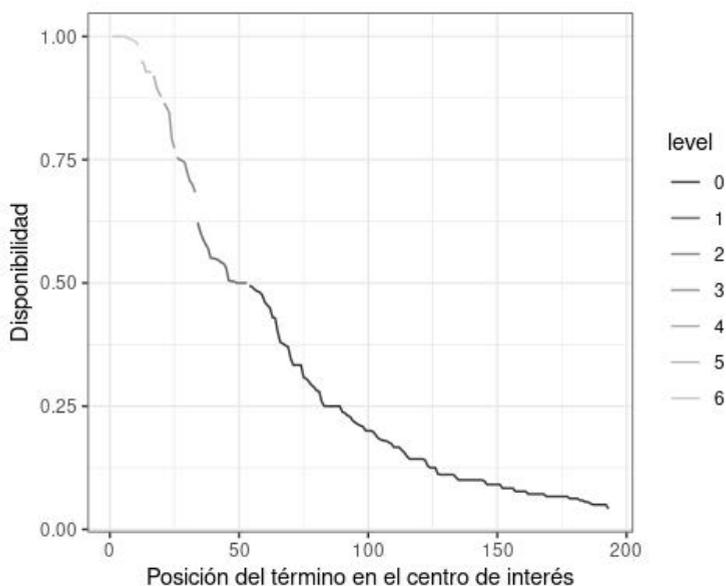
```
arrange(-availability) %>%
```

```
ggplot(aes(x=order, y=availability, color=level)) + geom_line() +
```

```
xlab("Posición del término en el centro de interés") +
```

```
ylab("Disponibilidad") +
```

```
theme_bw() + scale_colour_grey()
```



Se han construido utilidades para ayudar a la representación de estos conjuntos de corte. El siguiente ejemplo muestra los principales conjuntos de corte obtenidos en cuatro de los centros de interés usados en el trabajo de Ávila y Villena (eds.) (2010).

```

clasificacion <- build.availability.levels(levels)
clasificacion %>%
  filter(level> 0) %>% filter(centers %in% c("01","02","03","04")) %>%
  flextable() %>%
  set_header_labels(centers = "Centro de interés",
                    words = "Palabras",
                    level = "Nivel de disponibilidad",
                    count = "Recuento") %>%
  width(j=c(1,2,3), width=.75) %>%
  width(j=4, width=4) %>%
  align(i=4, align="right") %>%
  theme_booktabs()

```

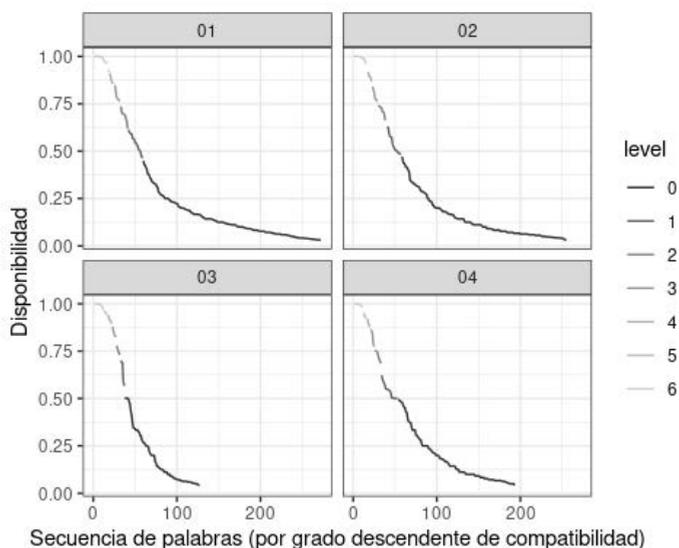
| Centro de interés | Nivel de disponibilidad | Recuento | Vocablos |
|-------------------|-------------------------|----------|--|
| 01 | 6 | 17 | cabeza, brazo, ojo, mano, pierna, pie, corazón, dedo, nariz, boca, pelo, oreja, hueso, hígado, músculo, tronco, uña |
| 01 | 5 | 2 | pulmón, riñón |
| 01 | 4 | 5 | estómago, extremidad, ceja, oído, órgano |
| 01 | 3 | 8 | cara, cerebro, codo, páncreas, rodilla, diente, sangre, hombro |
| 01 | 2 | 9 | cuello, esqueleto, pestaña, cintura, tibia, antebrazo, lengua, tobillo, vello |
| 01 | 1 | 18 | vena, enfermedad, espalda, pecho, labio, garganta, peroné, muñeca, bazo, esófago, culo, fémur, torso, salud, barriga, cadera, falange, piel |
| 02 | 6 | 16 | pantalón, camisa, falda, camiseta, calcetín, jersey, vestido, zapato, abrigo, chaqueta, blusa, braga, calzoncillo, sujetador, bufanda, corbata |
| 02 | 5 | 1 | (pantalón) vaquero |
| 02 | 4 | 5 | moda, tejido, media, sombrero, bañador |
| 02 | 3 | 7 | zapatilla, chándal, traje, camisón, anorak, bermudas, tanga |
| 02 | 2 | 10 | cinturón, bikini, rebeca, gorro, color, pijama, bolso, guante, necesario, chaleco |
| 02 | 1 | 18 | estilo, vestir, chaquetón, pañuelo, felpa, rebajas, diseño, comodidad, gabardina, botón, lana, (pantalón) pirata, falda vaquera, vestimenta, bota, sudadera, slip, algodón |
| 03 | 6 | 12 | cocina, salón, (cuarto de) baño, dormitorio, comedor, terraza, salita, entrada, (cuarto de) aseo, pasillo, lavadero, ventana |
| 03 | 5 | 4 | hall, recibidor, (cuarto) trastero, habitación |
| 03 | 4 | 6 | tejado, jardín, techo, escalera, sala de estar, vestidor |
| 03 | 3 | 5 | patio, garaje, pared, porche, balcón |
| 03 | 2 | 5 | despensa, buhardilla, sótano, suelo, vestíbulo |

| Centro de interés | Nivel de disponibilidad | Recuento | Vocablos |
|-------------------|-------------------------|----------|--|
| 03 | 1 | 5 | chimenea, puerta, despacho, azotea, estudio |
| 04 | 6 | 11 | mesa, silla, cama, sofá, sillón, armario, mesita de noche, aparador, cómoda, estantería, lámpara |
| 04 | 5 | 5 | ropero, mesa de comedor, espejo, vitrina, encimera |
| 04 | 4 | 4 | cuadro, lavadora, librería, frigorífico |
| 04 | 3 | 5 | tresillo, mueble de cocina, mueble-bar, lavavajillas, zapatero |
| 04 | 2 | 8 | butaca, tele(visión), mesilla de noche, escritorio, repisa, taquillón, tocador, cajonera sinfonier, hornilla, lámpara de techo, mueble de salón, alacena, mesa camilla, cama de matrimonio, mesa de tele(visión), taburete, comodín, mecedora, televisor, mesa de salón, ordenador, cortina, armario de niños, barbacoa, cenefa, mueble de entrada, rinconero |
| 04 | 1 | 20 | |

Tabla 6. Elementos seleccionados por niveles de corte en cinco centros de interés. Fuente: Ávila y Villena (eds.) (2010)

```

levels %>%
  mutate(level = factor(level)) %>%
  filter(centers %in% c("01","02","03","04")) %>%
  arrange(-availability) %>%
  ggplot(aes(x=order, y=availability, color=level)) + geom_line() + facet_wrap(~centers) +
  xlab("Secuencia de palabras (por grado descendente de compatibilidad)") +
  ylab("Disponibilidad") +
  theme_bw() + scale_colour_grey()
    
```



3.5. Utilidad *DispoCen*

Para facilitar el acceso a esta herramienta se ha dispuesto una sencilla utilidad gráfica que construye un documento básico con posibilidades que, en nuestra opinión, cubren las necesidades generales para el estudio de la disponibilidad y la centralidad. Este documento puede ser fácilmente modificable para adaptarlo a los requerimientos de cada usuario. Incluyendo la carga de la librería se puede actualizar como:

```
library(dispocen)
dispocen::runUtility()
```

Se abre entonces un interfaz gráfico con el siguiente aspecto:

centralex user interface

File to process

Choose file to load

Browse... No file selected

Model to apply

Model to apply

- Lopez-Strassburger
- Avila-Sanchez (Zipf + Additive)

 Generate report

Cuando carguemos un archivo de datos, solo tenemos que seleccionar el modelo de procesamiento (López-Strassburger o Ávila-Sánchez) y confirmar la operación. La utilidad calcula el proceso elegido y ofrece la descarga de un documento que ha de guardarse junto al archivo de datos. El almacenamiento de este documento junto al archivo de datos original es fundamental, pues de lo contrario, las utilidades posteriores dejarían de funcionar. Este documento incluye los bloques de código que llevan a cabo los procedimientos que hemos considerado útiles y que ofrecen sus resultados en el formato expuesto anteriormente. Muchos bloques están etiquetados con la opción *eval=FALSE*, que hace que no se ejecuten por defecto. El usuario interesado en la obtención de los resultados correspondientes, tan solo tendría que eliminar esa opción.

La utilidad proporciona las siguientes posibilidades:

1. Carga de datos y modelo de disponibilidad: el usuario dispondrá de una serie de bloques que le permitirán exponer los datos obtenidos por centro de interés, en forma de tabla y de gráfico, y un resumen gráfico de todos los centros de interés.
2. Niveles de centralidad/disponibilidad: con esta opción se obtiene una clasificación de los vocablos por niveles de centralidad y se construyen los listados de términos por nivel. También es posible obtener los listados con la clasificación y otros datos asociados junto a las tablas resumen. Se ha habilitado por defecto la presentación de los niveles de los términos por encima del nivel básico, de menor centralidad.

Nótese que este documento es un documento *RMarkdown* que es fácilmente modificable y personalizable. Se procesa entonces, dando lugar a un documento PDF, Word o HTML, entre otras opciones, que puede utilizarse como base para exportar los resultados.

3.6. Ejemplos de uso

Hemos decidido, de forma premeditada, no considerar en el análisis el uso de las primeras columnas de datos destinadas a la codificación de la información sociológica de los informantes porque, siguiendo los criterios de la normalización de bases de datos y de análisis de datos, su limitado formato y redundancia no constituye una buena práctica.

No obstante, presentamos a continuación algunas propuestas que consideramos mucho más productivas para llevar a cabo los análisis con filtros sociológicos.

3.6.1. Inclusión de los datos sociológicos

En primer lugar, es importante evitar la redundancia de los datos, tanto por la dificultad que implica mantener su coherencia cuando se encuentran duplicados en varios documentos, como para facilitar su actualización constante. Partimos, entonces, de una base de datos sociológica independiente donde cada sujeto aparece asociado a sus características personales. Nótese que esta base de datos puede contener tantas variables como el investigador crea conveniente y, además, dichas variables no estarían limitadas por un determinado número de posibles variantes, inconvenientes observados en programas anteriores.

```
bs <- read.csv("BaseSociologica.csv", colClasses = "character")
bs <- bs %>%
  mutate(EDAD = strtoi(EDAD),
         AÑOSDEESTUDIO = strtoi(AÑOSDEESTUDIO)) %>%
  rename(GENERO = SEXO)
bs %>%
  head() %>%
  select(SUJETO, GENERO, EDAD, AÑOSDEESTUDIO) %>%
  flextable() %>%
  autofit()
```

| SUJETO | GÉNERO | EDAD | AÑOS DE ESTUDIO |
|--------|--------|------|-----------------|
| 001 | 1 | 69 | 4 |
| 002 | 0 | 71 | 4 |
| 003 | 0 | 20 | 12 |
| 004 | 1 | 24 | 17 |
| 005 | 0 | 24 | 17 |
| 006 | 1 | 26 | 15 |

Tabla 7. Ejemplo de base de datos sociológica

3.6.2. Análisis por género

Si se desea llevar a cabo un análisis de centralidad léxica distinguiendo por géneros, simplemente habría que filtrar las realizaciones a procesar seleccionando a los sujetos que cumplen las características requeridas:

```
idHombres <- bs %>%
  filter(GENERO == "1") %>%
  select(SUJETO) %>%
  unlist()
idHombres

## SUJETO1 SUJETO2 SUJETO3 SUJETO4 SUJETO5 SUJETO6 SUJETO7 SUJETO8
## "001" "004" "006" "010" "012" "014" "015" "022"
## SUJETO9 SUJETO10 SUJETO11 SUJETO12 SUJETO13 SUJETO14 SUJETO15 SUJETO16
## "026" "029" "032" "035" "036" "039" "041" "043"
## SUJETO17 SUJETO18 SUJETO19 SUJETO20 SUJETO21 SUJETO22 SUJETO23 SUJETO24
## "046" "049" "050" "051" "052" "053" "055" "058"
## SUJETO25 SUJETO26 SUJETO27 SUJETO28 SUJETO29 SUJETO30 SUJETO31 SUJETO32
## "061" "064" "065" "070" "073" "075" "076" "097"
## SUJETO33 SUJETO34
## "098" "099"

idMujeres <- bs %>%
  filter(GENERO == "0") %>%
  select(SUJETO) %>%
  unlist()
idMujeres

## SUJETO1 SUJETO2 SUJETO3 SUJETO4 SUJETO5 SUJETO6 SUJETO7 SUJETO8
## "002" "003" "005" "008" "009" "013" "019" "021"
## SUJETO9 SUJETO10 SUJETO11 SUJETO12 SUJETO13 SUJETO14 SUJETO15 SUJETO16
## "023" "024" "025" "027" "028" "030" "031" "033"
## SUJETO17 SUJETO18 SUJETO19 SUJETO20 SUJETO21 SUJETO22 SUJETO23 SUJETO24
## "034" "037" "038" "040" "042" "045" "047" "048"
## SUJETO25 SUJETO26 SUJETO27 SUJETO28 SUJETO29 SUJETO30 SUJETO31 SUJETO32
## "056" "059" "060" "062" "063" "066" "067" "068"
## SUJETO33 SUJETO34 SUJETO35 SUJETO36 SUJETO37 SUJETO38
## "069" "071" "072" "074" "077" "096"
```

Entonces, podemos separar las realizaciones, si los sujetos pertenecen al grupo de hombres o mujeres:

```
data %>%
  filter(users %in% idMujeres) %>%
  head()
##   infos users centers words
## 1 12131 002    01 riñón, c...
## 2 12213 003    01 brazo, m...
## 3 12214 005    01 cabeza, ...
## 4 11232 008    01 cabeza, ...
```

```
## 5 11221 009 01 músculo,....
## 6 12224 013 01 cabeza, ....
data %>%
  filter(users %in% idHombres) %>%
  head()
##   infos users centers words
## 1 21131 001 01 mano, pi....
## 2 22214 004 01 brazo, o....
## 3 22213 006 01 pie, man....
## 4 22224 010 01 esquelet....
## 5 22224 012 01 cabeza, ....
## 6 22212 014 01 mano, pi....
```

A partir de este momento, se pueden procesar de la forma habitual:

```
disponibilidadHombres <-
  build.avilasanchez.availability(data %>%
    filter(users %in% idHombres))
head(disponibilidadHombres) %>%
  select(centers, words, availability) %>%
  flextable() %>%
  autofit()
```

| Centros | Palabras | Disponibilidad |
|---------|----------|----------------|
| 01 | cabeza | 0.9999756 |
| 01 | mano | 0.9987419 |
| 01 | brazo | 0.9984684 |
| 01 | ojo | 0.9978411 |
| 01 | pierna | 0.9907053 |
| 01 | pie | 0.9894805 |

Tabla 8. Muestra de términos más disponibles para los hombres.
Centro de interés 1. Fuente: Ávila y Villena (eds.) (2010)

```
disponibilidadMujeres <-
  build.avilasanchez.availability(data %>%
    filter(users %in% idMujeres))
head(disponibilidadMujeres) %>%
  select(centers, words, availability) %>%
  flextable() %>%
  autofit()
```

| Centros | Palabras | Disponibilidad |
|---------|----------|----------------|
| 01 | brazo | 0.9999198 |
| 01 | ojo | 0.9998306 |
| 01 | cabeza | 0.9986111 |
| 01 | mano | 0.9970720 |
| 01 | pierna | 0.9963897 |
| 01 | pie | 0.9926613 |

Tabla 9. Muestra de términos más disponibles para las mujeres.
Centro de interés 1. Fuente: Ávila y Villena (eds.) (2010)

Esta información se puede combinar para realizar análisis comparativos, enlazando los datos por centros de interés y términos, por ejemplo:

```
dispComp <-
  inner_join(disponibilidadHombres %>%
    select(centers, words, availability) %>%
    rename(avHombres = availability),
    disponibilidadMujeres %>%
    select(centers, words, availability) %>%
    rename(avMujeres = availability),
    by = c("centers", "words"))
dispComp %>%
  head() %>%
  flextable() %>%
  autofit()
```

| Centros | Palabras | Disponibilidad Hombres | Disponibilidad Mujeres |
|---------|----------|---------------------------|---------------------------|
| 01 | cabeza | 0.9999756 | 0.9986111 |
| 01 | mano | 0.9987419 | 0.9970720 |
| 01 | brazo | 0.9984684 | 0.9999198 |
| 01 | ojo | 0.9978411 | 0.9998306 |
| 01 | pierna | 0.9907053 | 0.9963897 |
| 01 | pie | 0.9894805 | 0.9926613 |

Tabla 10. Muestra de análisis comparativos por géneros. Centro de interés 1.
Fuente: Ávila y Villena (eds.) (2010)

```
dispComp %>%
```

```
  ggplot(aes(x=avHombres, y=avMujeres, color=centers)) + geom_point() +  
  theme_bw() + scale_colour_grey()
```

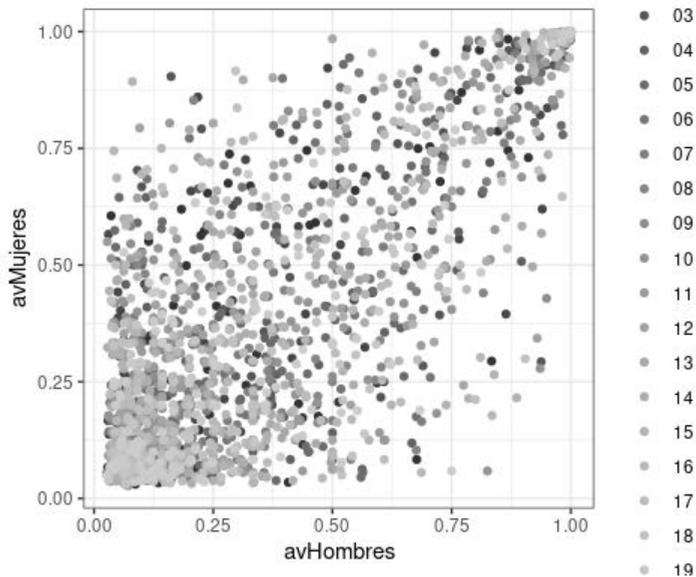


Gráfico 5. Comparación de términos según su índice de disponibilidad por géneros.
Centro de interés 1. Fuente: Ávila y Villena (eds.) (2010)

El Gráfico 5 nos permite observar que los términos muy disponibles y poco disponibles tienden a ser similares en ambos grupos. Sin embargo, los términos que aparecen en los diversos niveles intermedios parecen variar de forma muy significativa.

La Tabla 11 muestra la distribución de vocablos por niveles en ambos grupos de los primeros centros de interés usados en el trabajo de Villena Ávila (eds.) (2010):

```
levelsHombres <-  
  classify.availability.levels(disponibilidadHombres)  
clasificacionHombres <- build.availability.levels(levelsHombres)  
  
levelsMujeres <-  
  classify.availability.levels(disponibilidadMujeres)  
clasificacionMujeres <- build.availability.levels(levelsMujeres)  
  
clasificacionGenero <-  
  inner_join(clasificacionHombres %>%  
    select(centers, level, words) %>%  
    rename(wordsHombres = words),  
    clasificacionMujeres %>%  
    select(centers, level, words) %>%
```

```

rename(wordsMujeres = words),
by=c("centers","level"))

clasificacionGenero %>%
  filter(level> 0) %>%
  filter(centers %in% c("01", "02", "03", "04")) %>%
  flextable() %>%
set_header_labels(centers = "Centro de interés",
  wordsHombres = "Palabras hombres",
  wordsMujeres = "Palabras mujeres",
  level = "Nivel de disponibilidad",
  count = "Recuento") %>%
  width(j=c(1,2),width=.75) %>%
width(j=c(3,4),width=2.5) %>%
  align(j=c(3,4),align="right") %>%
theme_booktabs()

```

| Centro de interés | Nivel de disponibilidad | Palabras hombres | Palabras mujeres |
|-------------------|-------------------------|---|---|
| 01 | 6 | cabeza, mano, brazo, ojo, pierna, pie, corazón | brazo, ojo, cabeza, mano, pierna, pie, nariz, dedo, corazón, pelo, boca |
| 01 | 5 | hueso | |
| 01 | 4 | dedo, oreja, boca, tronco, nariz | uña, oreja, hígado |
| 01 | 3 | músculo, extremidad, pelo, hígado | músculo, pulmón, riñón, oído, tronco |
| 01 | 2 | estómago, pulmón, riñón, cerebro, esqueleto, uña, órgano | codo, hueso, ceja, estómago, hombro, rodilla, cara, cuello, antebrazo, vello |
| 01 | 1 | tibia, sangre, ceja, cara, vena, pestaña, peroné, páncreas, diente, oído, barriga | páncreas, diente, lengua, órgano, cintura, tobillo, enfermedad, cerebro, espalda, labio, sangre |
| 02 | 6 | pantalón, falda, camisa, vestido, camiseta, jersey, calcetín, blusa | camisa, pantalón, camiseta, calcetín, falda, zapato, jersey, chaqueta, calzoncillo |
| 02 | 5 | abrigo, sujetador, braga | vestido, abrigo, corbata |
| 02 | 4 | zapato, moda | sombrero |
| 02 | 3 | chaqueta, tejido, bufanda | bufanda, blusa, braga, bañador |
| 02 | 2 | (pantalón) vaquero, media, calzoncillo, estilo, bolso, bermudas | (pantalón) vaquero, traje, zapatilla, chándal, gorro, rebeca, vestir |

| Centro de interés | Nivel de disponibilidad | Palabras hombres | Palabras mujeres |
|-------------------|-------------------------|---|---|
| 02 | 1 | pijama, biquini, camisón, zapatilla, comodidad, falda vaquera, vestimenta, corbata, chándal, diseño, (pantalón) pirata, chaleco, chaquetón | anorak, sujetador, media, tanga, cinturón, guante, felpa, tejido, color, camisón, rebajas, necesario |
| 03 | 6 | cocina, salón, (cuarto de) baño, dormitorio, terraza, comedor, salita | cocina, salón, dormitorio, (cuarto de) baño, comedor, terraza, entrada |
| 03 | 5 | recibidor | (cuarto de) aseo, salita |
| 03 | 4 | pasillo, (cuarto de) aseo | techo, pasillo, hall |
| 03 | 3 | lavadero, entrada, (cuarto) trastero, jardín | lavadero, ventana, tejado |
| 03 | 2 | ventana, sala de estar, habitación, porche, vestidor, despensa | pared, habitación, escalera, (cuarto) trastero, garaje, vestíbulo, patio |
| 03 | 1 | hall, escalera, patio, despacho, tejado | balcón, chimenea, suelo, vestidor, jardín, sala de estar, sótano |
| 04 | 6 | mesa, cama, silla, sofá, sillón, armario | silla, mesa, sofá, sillón, cama, aparador, armario, mesita de noche |
| 04 | 5 | | cómoda |
| 04 | 4 | estantería, ropero, cómoda, lámpara, mesita de noche | estantería |
| 04 | 3 | aparador, lavadora, frigorífico | cuadro, lámpara, encimera, mesa de comedor, librería, espejo |
| 04 | 2 | vitrina, mesa de comedor, taquillón, zapatero, espejo, mueble de cocina, mesilla de noche | tresillo, vitrina, mueble-bar, cajonera, butaca |
| 04 | 1 | encimera, lámpara de techo, escritorio, tocador, mueble-bar, tele(visión), alacena, armario de niños, barbacoa, mueble de entrada, lavavajillas, cuadro, repisa | mueble de cocina, lavavajillas, lavadora, televisor, ropero, mecedora, cama de matrimonio, cenefa, comodín, rinconero, tele(visión) |

Tabla 11. Distribución de vocablos por niveles en 4 centros de interés. Fuente: Ávila y Villena (eds.) (2010)

Mediante un diagrama de puntos, en el que cada eje representa las disponibilidades obtenidas para hombres y mujeres de aquellos términos que no se han clasificado en el mismo nivel, obtenemos el resultado que muestra el Gráfico 6 clasificado según la diferencia de nivel de los términos:

```

levelsGenero <-
  inner_join(levelsHombres %>%
    select(centers, level, words, availability) %>%
    rename(avHombre = availability, levelHombre=level),
    levelsMujeres %>%
    select(centers, level, words, availability) %>%
    rename(avMujer = availability, levelMujer=level),
    by=c("centers", "words"))
levelsGenero %>%
  filter(levelHombre != levelMujer) %>%
  mutate(diffLevel = factor(abs(levelHombre - levelMujer))) %>%
  ggplot(aes(x=avHombre, y=avMujer, color=diffLevel)) + geom_point() +
  theme_bw() + scale_colour_grey()

```

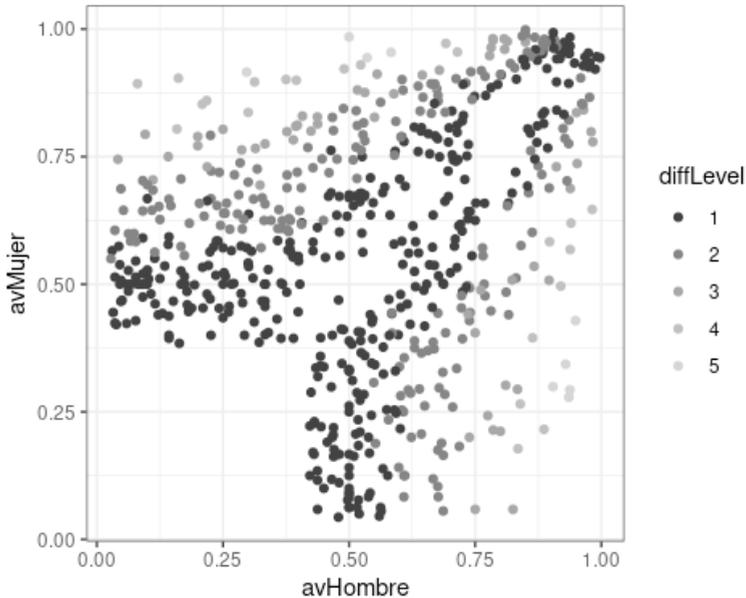


Gráfico 6. Términos clasificados en diferente nivel para mujeres y hombres en todos los centros de interés. Fuente: Ávila y Villena (eds.) (2010)

El Gráfico 6 nos permite observar que en la mayor parte de los casos en los que los vocablos se clasifican en niveles distintos, las diferencias son escasas.

```

diffs <- tapply(levelsGenero$levelHombre == levelsGenero$levelMujer,
  levelsGenero$centers, mean) * 100

```

```

diffs
##    01    02    03    04    05    06    07    08
## 71.66667 57.31707 60.00000 66.27907 66.20690 61.64384 71.42857 64.83516

```

```
##      09      10      11      12      13      14      15      16
## 66.66667 65.57377 71.42857 74.68354 65.15152 69.11765 64.95726 73.68421
##      17      18      19      20
## 71.30435      66.00000      73.77049      67.17557
summary(diffs)
##  Min.    1st Qu.    Median    Mean    3rd Qu.    Max.
## 57.32    65.10     66.47     67.44     71.43     74.68
```

De hecho, se puede observar que el porcentaje de términos clasificados en el mismo nivel en cada centro de interés oscila entre el 57 y el 74 por ciento.

4. CONCLUSIONES

DispoCen es una utilidad para el cálculo de la disponibilidad y la centralidad léxica. Pero sus posibilidades van mucho más allá de las utilidades señaladas. Al estar generada con una librería de herramientas en *R* por medio de programación funcional, sus aplicaciones son más abiertas y extensas. De momento, el investigador de la disponibilidad y la centralidad léxica encontrará en *DispoCen* una herramienta actualizada que sustituya, quizás, a programas obsoletos o de acceso restringido comunes hasta ahora para esta función.

En este documento hemos presentado cada uno de los códigos necesarios para la realización de los análisis que conducen a la obtención de las listas de disponibilidad y centralidad, con lo cual se abre a los usuarios de *DispoCen* una serie de posibilidades que hasta ahora eran inaccesibles: desde la verificación de los pasos realizados hasta el desarrollo y adaptación a cada situación particular de análisis gracias a la replicabilidad que ofrece esta aproximación. Aunque la carga de datos replica el formato habitual en la preparación de los archivos para el cálculo de la disponibilidad, en el interfaz gráfico hemos prescindido de la posibilidad de calcular los índices propuestos discriminando por grupos sociológicos excesivamente restringidos. La redundancia de esta información sugiere la necesidad de separar los datos sociológicos y léxicos en dos archivos diferentes y realizar las operaciones donde se incluyan discriminaciones sociológicas mediante el empleo de códigos simples de filtrado y combinación de variantes. De esta manera, se multiplican las posibilidades de análisis, ya que la base de datos sociolingüística es abierta y puede contener tantas variables y variantes como el investigador desee considerar.

Actualmente está en preparación otro trabajo en el que se discuten en profundidad los modelos posibles y las diferentes opciones de aplicación, así como un estudio de diversos aspectos referidos a las bases teóricas del cálculo de la disponibilidad y su interpretación. Sin embargo, el desarrollo de estas herramientas posteriores requiere cierta destreza en el manejo de la sintaxis de *R*, así como el dominio de los conceptos de programación funcional, con lo que su presentación requiere otro ámbito de exposición.

REFERENCIAS BIBLIOGRÁFICAS

- Ávila Muñoz, A. M. y Sánchez Sáez, J. M. 2014. "Fuzzy sets and Prototype Theory: Representational model of cognitive community structures based on lexical availability trials, *Review of Cognitive Linguistics*, 12/1, pp. 133-159.
- Ávila Muñoz, A. M. y Villena Ponsoda, J. A. (eds). 2010. *Variación social del léxico disponible en la ciudad de Málaga. Diccionario y análisis*. Málaga: Editorial Sarriá.
- Bartol Hernández, J. A. 2001. "Reflexiones sobre la disponibilidad léxica". En Bartol, J. A. (coord.). *Nuevas aportaciones al estudio de la Lengua Española*. Salamanca: Luso-Española de Ediciones, pp. 221-235
- Callealta Barroso, F. J. y Gallego Gallego, D. J. 2016. "Medidas de disponibilidad léxica: comparabilidad y normalización", *Boletín de Filología*, LI, 1, pp. 39-92.
- Carcedo González, A. 2001. *Léxico disponible de Asturias*. Turku: Departamento de Español de la Universidad de Turku. Carnap.
- Moreno Fernández, F.; Moreno Fernández, J. E. y García de las Heras, A. J. 1995. "Cálculo de disponibilidad léxica. El programa LexiDisp", *Lingüística*, 7, pp. 243-250.
- Samper Padilla, José Antonio. 1999. "Léxico disponible y variación dialectal: datos de Puerto Rico y Gran Canaria". En Morales, Amparo *et al.* (eds.). *Estudios de lingüística hispánica. Homenaje a María Vaquero*. San Juan de Puerto Rico: Universidad, pp. 550-573.