

**PATRONES SOCIOLINGÜÍSTICOS DEL  
VOCABULARIO DISPONIBLE.  
CONDICIONAMIENTO ESTRATIFICACIONAL DE LA  
CAPACIDAD LÉXICA EN LA CIUDAD DE MÁLAGA.  
PROYECTO CONSOLEX**

Juan Andrés Villena Ponsoda  
Antonio Manuel Ávila Muñoz  
José María Sánchez Sáez  
Universidad de Málaga

## 1 INTRODUCCIÓN

### *1.1 LA DISPONIBILIDAD LÉXICA EN EL PROYECTO CONSOLEX*

En este trabajo se avanzan los resultados más importantes del Proyecto de estudio de los Condicionantes Sociales del Léxico Disponible en la ciudad de Málaga. Proyecto de Investigación de Excelencia. Consejería de Ciencia e Innovación de la Junta de Andalucía. HUM-315, 2006-2009 (Ávila y Villena (eds.) 2010; Ávila y Sánchez 2011).

Gracias a los modelos de representación del vocabulario virtual basados en la teoría de los conjuntos difusos, el estudio de la situación de los hablantes en el

continuo de centralización y descentralización en un centro de interés determinado permite calcular índices precisos de competencia léxica individual y aprovechar los resultados de las pruebas clásicas de disponibilidad léxica.

El efecto directo de las variables de posición social (especialmente el prestigio ocupacional y los estudios del individuo) en la capacidad léxica del hablante es, como se sabe, relevante. Sin embargo, la biografía social de los individuos establece vías de acceso a los mercados sociales y lingüísticos no directamente relacionadas con los conocimientos adquiridos en el período de educación formal. La capacidad de usar y reconocer léxico especializado actúa como una variable de intervención entre la educación del hablante y el tamaño y originalidad del vocabulario aportado por este en las pruebas de disponibilidad.

### 1.1.1 INNOVACIÓN Y TRADICIÓN

Se desarrolla el estudio de la disponibilidad léxica en una ciudad completa. Se han aprovechado los conocimientos, los métodos de trabajo de campo y las herramientas analíticas surgidas y perfeccionadas en la investigación de sectores parciales de las poblaciones, en especial de jóvenes preuniversitarios (Gougenheim *et alii* 1953, 1964), aunque el campo de nuestro estudio puede considerarse relativamente nuevo: la ciudad completa. Hasta el momento, que sepamos, no se ha considerado de interés extender este tipo de investigaciones al ámbito general de una ciudad.

### 1.1.2 REVISIÓN DE LOS MODELOS MATEMÁTICOS

Se han repasado las bases de las fórmulas utilizadas para la obtención de los listados de palabras (Butrón 1987; López Chávez y Strassburger 1987; López Morales 1999; Lorán 1983; Lorán y López Morales 1983) con la finalidad de explicar el proceso objeto de estudio y, por extensión, reutilizar la información obtenida para construir nuevos aspectos cuantificables. Los procedimientos matemáticos empleados hasta la fecha han resultado ser útiles para la obtención de

las listas de palabras. Pero estos, por sí mismos, no dejan de proporcionar simples recuentos descriptivos sin capacidad explicativa ni predictiva, por lo que hemos considerado oportuno ampliar el enfoque.

### 1.1.3 EL INTERÉS POR LOS HABLANTES

Uno de los principales objetivos del estudio consistió en la redefinición de la disponibilidad como variable a disposición de los hablantes individuales (y no tanto como propiedad de los elementos léxicos). Nuestra propuesta se basa en una aproximación matemática pensada para desentrañar la estructura de acceso al vocabulario por parte de los hablantes que nos ofrezca una imagen de su capacidad léxica. Al disponer de información sociorreticular de los informantes, ha sido posible confirmar algunas hipótesis sobre la determinación que la posición social y otras variables de hablante ejercen sobre el vocabulario virtual de los individuos.

## 1.2 OBJETIVOS

### 1.2.1 DICCIONARIO DE DISPONIBILIDAD LÉXICA DE LA CIUDAD DE MÁLAGA

Un diccionario de disponibilidad comunitario puede acrecentar, si cabe, el interés asociado a los datos procedentes del sector de los estudiantes preuniversitarios. En particular, se trata de la comparación de los resultados de la disponibilidad léxica en la ciudad con los obtenidos de los grupos de adolescentes (Ávila 2006). Se plantean cuestiones clásicas sobre la riqueza/pobreza, el déficit o el desinterés de los sectores poblacionales más jóvenes. Las comparaciones entre los inventarios de los jóvenes preuniversitarios y los de los demás grupos sociales de la ciudad producen contrastes que son síntomas, probablemente, de mecanismos y problemas más profundos en la sociedad.

### 1.2.2 ANÁLISIS DE LA VARIACIÓN SOCIAL DEL LÉXICO DISPONIBLE

Se trata de un desarrollo de los trabajos anteriores, tanto en lo tocante a la extensión del universo de estudio, como en el número y la variación de los datos sobre los informantes. El manejo de una información considerable sobre los hablantes y sobre su organización social estructural y reticular permite disponer de resultados notablemente ricos. En particular, los efectos de las variables de posición social sobre el tamaño y la originalidad de las listas de disponibilidad constituyen aspectos de interés cuyo conocimiento es previo a cualquier intento de planificación en la escuela y en la sociedad.

### 1.3 MUESTRA

La muestra de hablantes ( $n = 72$ ) con afijación uniforme se estratificó por edad, sexo y nivel educacional. Se utilizó el modelo utilizado en trabajos anteriores sobre la misma comunidad de habla (Tabla 1).

**Tabla 1. Distribución de la muestra de hablantes**

Educación	Hombres				Mujeres			
	Prim.	Secund.	Univ.	Total	Prim.	Secund.	Univ.	Total
<b>Edad</b>								
<b>20-34</b>	4	4	4	12	4	4	4	12
<b>35-54</b>	4	4	4	12	4	4	4	12
<b>54</b>	4	4	4	12	4	4	4	12
<b>Total</b>	12	12	12	36	12	12	12	36

### 1.4 METODOLOGÍA

Junto a las tradicionales pruebas léxicas escritas de dos minutos de duración para cada uno de los veinte centros de interés propuestos, se pasó asimismo a los informantes un cuestionario de palabras técnicas o especializadas relacionadas con los veinte centros de interés (dos palabras por cada CI). La tarea consistía en que los sujetos relacionaran cada una de las piezas léxicas con un centro de interés determinado. Se rellenó asimismo un cuestionario sociorreticular con cada

informante. Se tuvo en cuenta un considerable número de variables de postestratificación. La hipótesis fue que algunas tendrían valor explicativo sobre la variación léxica. La información sociológica se recogió por medio de un cuestionario sociorreticular específico que los informantes rellenaban tras la prueba léxica.

### 1.5 HIPÓTESIS

#### 1.5.1 COMPETENCIA LÉXICA DEL HABLANTE

Se ha pretendido calcular la competencia léxica individual a partir de los datos recogidos en las pruebas estándar de disponibilidad. El cálculo se efectúa mediante la construcción de modelos basados en conjuntos difusos, que representan los marcos o centros de interés y que permiten obtener puntuaciones individuales de centralización en dichos marcos, en función de las coincidencias y discrepancias entre las listas individuales (Ávila y Sánchez 2010, Villena, Ávila y Sánchez 2012).

#### 1.5.2 ESTRATIFICACIÓN SOCIAL DEL LÉXICO DISPONIBLE

Las diferencias en la competencia léxica están determinadas por la posición social de los individuos. Los años de estudio cursados condicionan el tamaño y la descentralización del vocabulario virtual de los hablantes. Dada la etapa a la que se produce la formación reglada, se supone que los conocimientos técnicos adquiridos durante el período posterior y durante toda la vida acrecientan el efecto de la educación. Por esta razón, junto a los indicadores de posición (educación, prestigio ocupacional, ingresos) se utiliza la variable de intervención (*Pident*) mencionada más arriba, que mide esta última capacidad (capacidad léxica especializada). Por último, aunque la edad del hablante no influye de manera directa en la capacidad léxica, sin embargo al interaccionar con la educación

ofrece interesantes perspectivas que rechazan la idea preconcebida de una relación lineal entre edad y capacidad léxica (Villena y Ávila 2010).

## 2 DESCENTRALIZACIÓN Y CAPACIDAD LÉXICA

### 2.1 HIPÓTESIS

1. Centro de interés. Un centro de interés (CI) es un punto de acceso a una red de elementos léxicos que el hablante relaciona entre sí y con el concepto enunciado como tal CI.

2. Estructura del CI. La estructura del centro de interés es subjetiva e inherente al hablante y, por tanto, no es estrictamente extrapolable a otros hablantes.

3. Centro de interés compartido. Las redes léxicas de un conjunto de individuos relacionados tienden a tener formas similares, debido a condicionantes sociológicos y culturales. Esta metaestructura compartida es lo que denominamos centro de interés y es el objetivo de consolidación para nuestro estudio.

4. Pruebas de disponibilidad. En las pruebas de obtención del léxico: (1) se accede a un elemento léxico cercano al concepto propuesto; (2) al ir consumiendo piezas léxicas, se produce un alejamiento progresivo, mediante elementos relacionados, del concepto central; (3) cuando el hablante considera que se ha alejado demasiado de dicho concepto central, vuelve a escoger un término cercano al prototipo, produciéndose así un proceso de reentrada que hace que cada muestra individual no sea estrictamente relevante de la estructura del léxico del individuo; (4) son determinantes en el proceso tanto la capacidad léxica de cada hablante (el tamaño y la originalidad de su vocabulario) como su velocidad de acceso; sin embargo, en las pruebas clásicas de disponibilidad no se controla la velocidad de acceso al léxico; (5) el orden de aparición de las palabras es relevante con respecto a su accesibilidad.

5. Estructura del léxico. Se considera que la estructura del léxico de un centro de interés está compuesta de: (1) el núcleo o conjunto de términos básicos

accesibles a todos los hablantes; (2) la periferia o conjunto de términos no accesibles a todos los hablantes; (3) el paso del núcleo a la periferia no es brusco, sino gradual, de modo que estamos ante una gradación en lugar de ante una clasificación dicotómica.

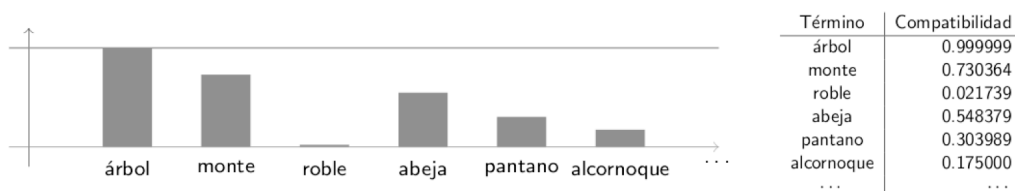
## 2.2 TEORÍA DE LOS CONJUNTOS DIFUSOS

### 2.2.1 COMPATIBILIDAD CON UN CONJUNTO

Esta teoría permite reinterpretar la pertenencia de elementos a conjuntos como la compatibilidad de esos elementos con el concepto representado por el conjunto (Zadeh 1965, Zimmermann 2001), permitiendo una gradación en estos grados de compatibilidad, que varía entre compatibilidad absoluta (1) e incompatibilidad absoluta (0).

Así pues, se puede representar el índice de disponibilidad en un CI como la compatibilidad del léxico con el prototipo proporcionado por el CI (Gráfico 1).

**Gráfico 1. Estructura de los centros de interés como una gradación de compatibilidad**



La interpretación de los valores obtenidos y de su gradación es simple:

Cercano a uno: pertenece al núcleo del centro de interés y es accesible para todos los hablantes.

Cercano a cero: pertenece a la periferia y es accesible solo a una parte de los hablantes.

### 2.2.2 CONSTRUCCIÓN DEL ESPECTRO DEL LÉXICO

Se denomina espectro léxico al conjunto de elementos léxicos enunciados en un CI, con sus correspondientes grados de compatibilidad con este centro. La determinación de esta información es la resolución del problema de la obtención de la disponibilidad de cada elemento. Al considerar la disponibilidad de un elemento como un grado de compatibilidad, se traduce el problema a la evaluación de estos grados de pertenencia.

Basándonos en la hipótesis –al menos en una primera aproximación– de que el orden del listado de piezas léxicas proporcionado por cada hablante es relevante respecto a la disponibilidad de cada término, se construye, para cada listado de términos, una representación de la disponibilidad realizada siguiendo una ley descendente con la posición. Para ello hemos escogido una ley proporcional a la de Zipf-Mandelbrot, ley que postula que en un corpus de lenguaje natural la frecuencia de cualquier término es inversamente proporcional a su posición en la ordenación en frecuencia de aparición<sup>1</sup>, y que es de uso común en este tipo de trabajos:

$$m(t) \propto \frac{1}{k}$$

donde  $k$  es la posición del término en la lista enunciada por el hablante.

Así, por ej., de la lista de términos siguiente:

*(1) tierra, flor, río, árbol, animal, sol, aire fresco, paz, incendio, clorofila, oveja, burro, pan blanco, buen olor*

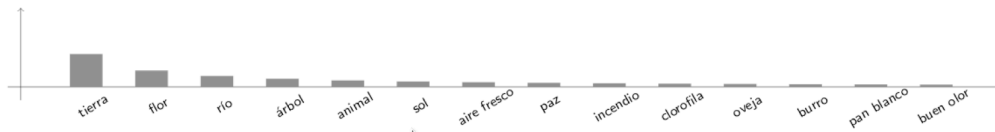
se obtendría una realización del espectro expresado por el individuo como la que se muestra en el Gráfico 2.

---

<sup>1</sup> En el sentido de que, en algún sentido, acumula información, no que la suma directamente.



**Gráfico 2. Representación del espectro léxico del CI 11. El campo**



### 2.2.3 INTEGRACIÓN DE LAS REALIZACIONES DE LOS HABLANTES

Partiendo de la hipótesis de que la estructura de un centro de interés es un constructo común de las estructuras léxicas de cada hablante, se integra la relevancia de cada término entre los distintos hablantes mediante una ley aditiva<sup>2</sup>. Hay multitud de operadores que cumplen las condiciones que necesitamos, pero ante la falta de información<sup>3</sup>, se ha optado por la más simple, que es la suma probabilística: si  $a$  y  $b$  son dos compatibilidades de un término para dos hablantes, la compatibilidad resultante se calcula como:

$$a + b - a \cdot b$$

que tiene como característica el que la aparición de un término siempre añade, en relación directa con la relevancia de su aparición, más relevancia al término. En el Gráfico 3, a continuación, se puede muestra una simulación de lo que ocurriría al integrar la información proporcionada por tres hablantes (tres iteraciones).

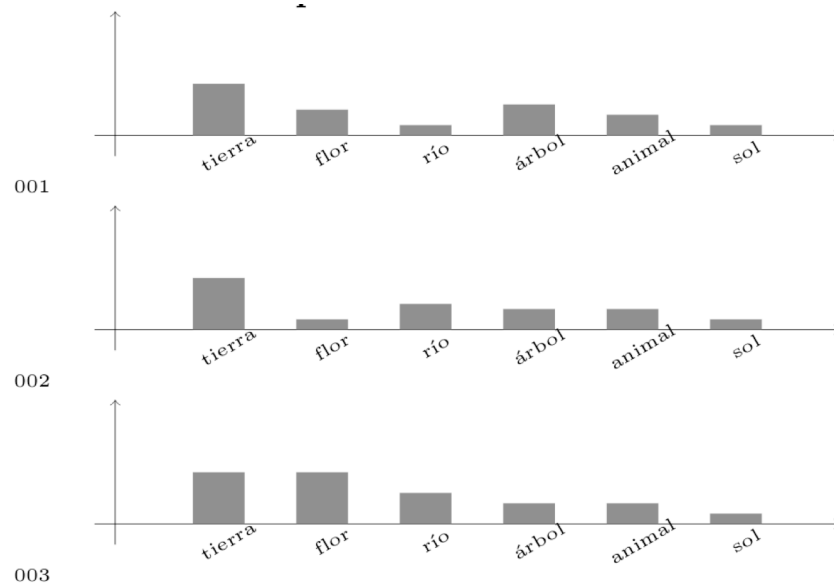
---

<sup>2</sup> Al menos hasta que se lleven a cabo las suficientes investigaciones que permitan determinar una opción más adecuada.

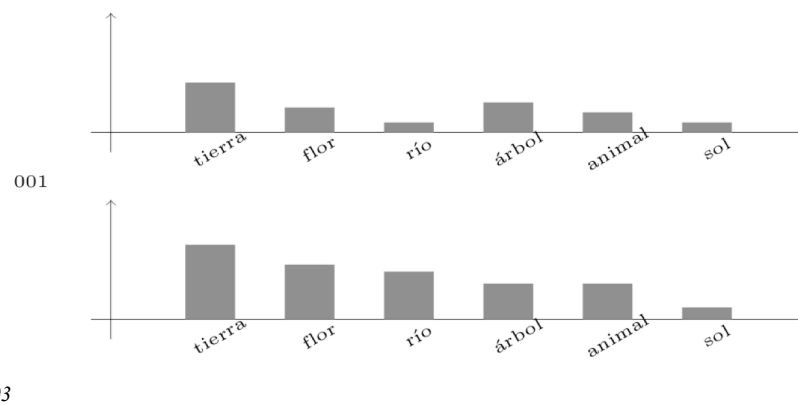
<sup>3</sup> Los términos se han reordenado para que estén en todas las figuras en el mismo orden. El orden de aparición en las listas ya está considerado en la evaluación de su compatibilidad.

**Gráfico 3. Proceso de integración de los espectros individuales en el CI 11. El campo**

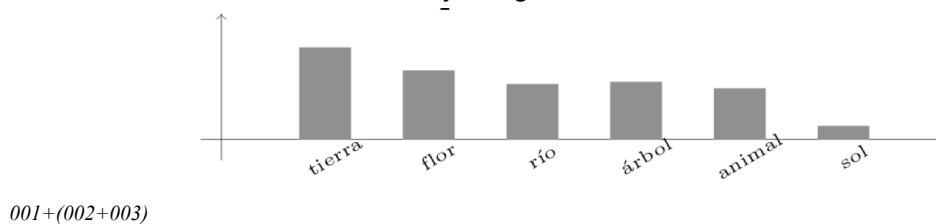
*Iteración 1:* se construyen las evaluaciones de las realizaciones de cada hablante (01, 02 y 03)<sup>4</sup>.



*Iteración 2:* se integra la información proporcionada por los hablantes 002 y 003



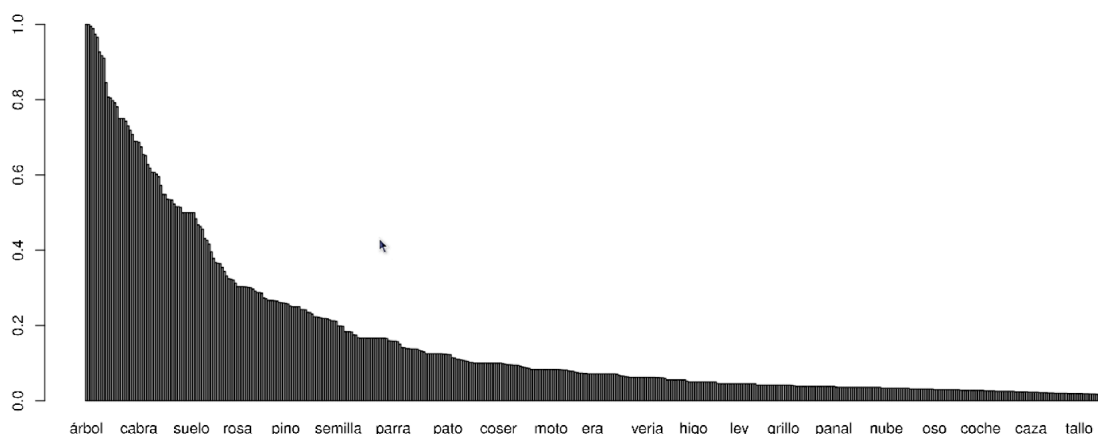
*Iteración 3:* se integra la información proporcionada por el hablante 001 a la información ya integrada



<sup>4</sup> Los términos se han reordenado para que estén en todas las figuras en el mismo orden. El orden de aparición en las listas ya está considerado en la evaluación de su compatibilidad.

Un ejemplo del resultado de este proceso en un caso real es el que se muestra en el Gráfico 4, más abajo, en el que los términos se han ordenado en función de su relevancia.

**Gráfico 4. Integración ordenada de espectros individuales en el CI 11. El Campo**



### 2.2.4 PROPIEDADES DE LA REPRESENTACIÓN DEL ESPECTRO

La principal característica del resultado obtenido es que si bien se obtienen así clasificaciones equivalentes a las logradas mediante otros modelos previos, el modelo de compatibilidad desarrollado es, en sí mismo, productivo; a saber:

Permite utilizar las herramientas que proporciona la teoría de conjuntos difusos.

Distingue el léxico entre CI diferentes o el mismo CI entre diferentes poblaciones, utilizando los operadores de diferencia y diferencia simétrica.

Determina el tamaño del léxico de un centro de interés: cardinalidad difusa.

Determina el valor de relevancia a partir del cual se puede considerar que un elemento léxico está en el núcleo de un centro de interés: Fuzzy Expected Value (FEV).

Al evaluarse en etapas, se pueden introducir otros tipos de valoraciones en cada evaluación individual, como el tiempo de reacción, que actualmente no se utiliza.

Por último, hace posible desarrollar construcciones hasta ahora no consideradas, como el cálculo de la capacidad léxica individual.

### 2.3 ÍNDICE DE CAPACIDAD LÉXICA INDIVIDUAL

#### 2.3.1 CONCEPTO DE CAPACIDAD LÉXICA INDIVIDUAL

Una de las hipótesis básicas de este trabajo es la de la construcción de una medida de la capacidad léxica individual a partir de la representación de la disponibilidad léxica. El índice de capacidad léxica individual debe indicar el volumen del vocabulario de un hablante y la velocidad a la que este accede a dicho vocabulario. A mayores volúmenes de vocabulario y a mayor velocidad de acceso, mayor capacidad léxica. Puesto que la velocidad no se ha obtenido en las pruebas, tan solo podemos trabajar con el volumen del léxico aportado.

#### 2.3.2 ELEMENTOS BÁSICOS

Los conceptos claves para nuestro desarrollo del concepto de ‘capacidad léxica individual’ son los siguientes:

1. Por hipótesis, el núcleo del CI estará disponible a *todos* los hablantes, aun cuando no lo hayan realizado.
2. El léxico diferenciador, al que no todos los hablantes tienen acceso, es el periférico.
3. Por lo tanto, el léxico más relevante de una alta capacidad léxica es aquel que está más alejado del núcleo.

4. La adici3n de nuevo léxico *nunca* supone una reducci3n de la capacidad léxica; esto es, siempre es acumulativo.

5. De aquí se obtiene, por tanto, que el concepto de especificidad es el opuesto al de disponibilidad del léxico.

6. Entonces, si la disponibilidad de un término en un centro de interés, como se construyó más arriba, es  $m_i(t)$ , entonces su especificidad es:

$$1 - m_i(t)$$

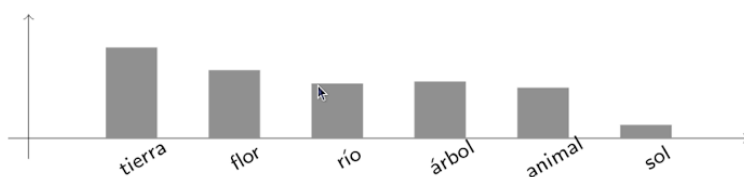
gracias a la interpretaci3n de la disponibilidad del léxico como compatibilidad con el concepto expresado por el centro de interés.

7. Utilizando una ley acumulativa, como la suma probabilística, se acumula la especificidad de los términos expuestos por el hablante.

### 2.3.3 SIMULACI3N

Considerando el proceso propuesto, se ha construido la simulaci3n que se muestra en el Gráfico 5.

**Gráfico 5. Simulaci3n de la construcci3n del espectro léxico individual (CI 11. El campo)**



Término	Compat.	Exclus.	Escalado	Acuml.
tierra	0.88	0.12	0.04	0.04
flor	0.66	0.34	0.11	0.15
río	0.53	0.47	0.16	0.28
árbol	0.55	0.45	0.15	0.39
animal	0.49	0.51	0.17	0.49
sol	0.13	0.87	0.29	0.64

Como puede observarse:

1. La primera columna corresponde al elemento léxico.
2. La segunda columna corresponde a su compatibilidad con el centro de interés del que se extrae la información.
3. La tercera columna corresponde a la exclusividad, que es el complementario a uno de la compatibilidad.
4. En la cuarta columna se ha llevado a cabo un escalado, para mantener la acumulación de los valores en un ámbito delimitado y manejable<sup>5</sup>
5. En la última columna se aplica la suma probabilística de cada valor escalado con la acumulación anterior.

Con todo ello se obtendría, como resultado final, que la capacidad léxica del individuo en cuestión se podría evaluar como 0.64. Con este procedimiento hemos trabajado en este estadio.

### *2.4 PROBLEMAS ABIERTOS*

En la actualidad, aunque creemos que el modelo expuesto tiene las suficientes ventajas como para considerar conveniente su presentación, también entendemos que quedan ámbitos sobre los que conviene seguir trabajando:

Los procesos de acumulación de información requieren un escalado previo de los datos, que en nuestro caso hemos llevado a cabo multiplicando por una cantidad fija, para que los datos obtenidos se sitúen en un rango manejable. Es necesario describir condiciones para determinar este proceso de escalado de forma objetiva<sup>6</sup>.

Se puede acumular más información a cada evaluación individual como, por ej., la evolución temporal. Las pruebas, tal y como están

---

<sup>5</sup> De hecho, en la actualidad estamos trabajando en una solución que consideramos muy prometedora.

<sup>6</sup> Véanse A. Llorente y J. Modéjar (1974), y A. Enrique Arias (2002) entre otros.

diseñadas en la actualidad, no consideran esta información, pero es posible introducirla. Sería una tarea interesante, y posiblemente ardua, determinar las formas de cuantificar esta información.

Ante la falta de modelos y de información hemos escogido las leyes y operadores más simples que reunieran las condiciones mínimas que consideramos importantes: la suma probabilística y la ley de Zipf-Mandelbrot. Sería conveniente discutir y probar diferentes opciones, y elegir aquellas que mejor se ajusten al problema en particular.

No existe un mecanismo objetivo para diferenciar elementos del léxico del CI poco frecuentes de elementos que no tienen ninguna relación con el centro de interés.

### 3 CONDICIONAMIENTO ESTRATIFICACIONAL. ANÁLISIS

#### 3.1 VARIABLES

##### 3.1.1 VARIABLE DEPENDIENTE

Índice de Descentralización Léxica (*ICoDD*), que se considera una representación de la competencia léxica individual medida a través de la prueba de disponibilidad descrita arriba.

##### 3.1.2 VARIABLES INDEPENDIENTES

Las variables de hablante consideradas (Tabla 2) fueron los indicadores de estatus (estudios, prestigio ocupacional, ingresos, etc.), los marcadores de red social (número, fortaleza de los lazos, etc.) y las variables de atributos personales (edad, sexo, procedencia). Se tuvo en cuenta asimismo una variable léxica de intervención: la capacidad léxica especializada (*Pident*). Esta se obtuvo, como vimos en 1.4, mediante una prueba complementaria paralela al test de disponibilidad y consistió en la ubicación correcta por parte de los informantes de dos piezas léxicas especializadas por cada centro de interés. Como se verá después, esta variable multiplica el efecto de algunas variables de hablante en la predicción de la capacidad léxica individual.



Tabla 2. Características de los hablantes y diferencias de sexo

	Total			Hombres			Mujeres		
	Media	Desv.	N	Media	Desv.	N	Media	Desv.	N
<b>Variables sociodemográficas</b>									
<b>Edad</b>	43.9	16.8	72	42.9	17.2	36	45.0	16.5	36
<b>Estudios</b>	11.2	5.0	72	11.3	5.1	36	11.0	5.0	36
<b>Estudios del padre</b>	7.8	5.8	66	8.8	6.2	34	6.7	5.2	32
<b>Estudios de la madre</b>	6.3	5.0	69	7.0	5.2	36	5.5	4.6	33
<b>EMedia(0-20)</b>	14.1	4.6	69	15.0	4.0	35	13.1	5.1	34
<b>Internet(horas)</b>	4.4	6.8	54	5.2	6.4	30	3.5	7.3	24
<b>Variables reticulares</b>									
<b>Tamaño</b>	9.6	3.4	72	9.5	3.0	36	9.7	3.8	36
<b>Contactos externos</b>	3.8	2.4	72	4.2	2.3	36	3.4	2.3	36
<b>Intensidad reticular</b>	6.1	3.1	72	6.0	2.9	36	6.1	3.3	36
<b>Multiplicidad</b>	8.6	17.4	72	9.0	16.4	36	8.2	18.5	36
<b>Variables lingüísticas</b>									
<b>ICoDD</b>	0.61	0.17	72	0.63	0.15	36	0.58	0.18	36
<b>Pident</b>	0.51	0.23	72	0.54	0.23	36	0.47	0.23	36

### 3.2 ANÁLISIS BIVARIANTE

#### 3.2.1 POSICIÓN SOCIAL

La capacidad léxica individual (ICoDD) depende del número de años de estudio cursados por el hablante (Tabla 3). Sin embargo, el porcentaje de variación explicado no es tan grande como se podría suponer ( $R^2 = 51.0\%$ ).

Tabla 3. Correlación entre las variables léxicas y los estudios y la edad (n = 72)

	<b>Pident</b>	<b>Estudios</b>	<b>Edad</b>
<b>Estudios</b>	0.812		
<b>Edad</b>	0.211	0.036	
<b>ICoDD</b>	0.719	0.645	-0.027

Correlaciones significativas ( $p < 0.01$ ) en negrita ( $p = 0.000$ )

La comunidad estudiada tiende a la bipolarización del vocabulario en función del estatus: (1) hablantes con estudios básicos y ocupación poco prestigiosa: vocabulario disponible poco especializado y nuclear, compartido por todos los hablantes y que se repite en todas las listas; (2) hablantes con estudios secundarios y universitarios con ocupaciones manuales especializadas o no manuales: vocabulario virtual menos nuclear y más especializado (Tabla 4).

**Tabla 4. Efectos de la educación en la descentralización léxica**

Grupos	Media	Desv.	n
Primarios	0.49	0.14	21
Resto	0.69	0.13	44

$t = -5.715(63)$ ; sig. 0.000

### 3.2.2 EDAD

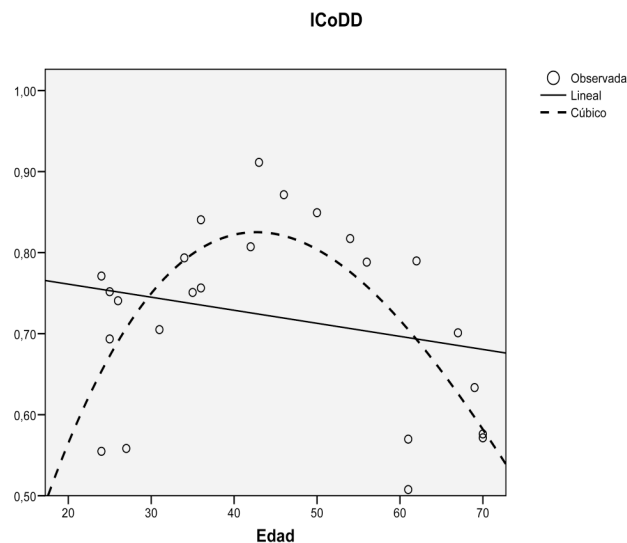
El efecto de la edad en la descentralización léxica no es lineal ni significativo en el sentido esperado, lo que tiende a rechazar la idea de sentido común sobre la incompetencia léxica de los jóvenes contemporáneos. La edad por sí misma no explica linealmente las diferencias del vocabulario virtual de los hablantes.

Existe, por el contrario, una leve correlación no lineal entre ambas variables. Se trata de una correlación curvilínea que implica una relación positiva en la primera mitad de los valores de edad (de 25 a 50 años, aproximadamente) y, a continuación, una correlación negativa (de 50 años en adelante). Es cierto, sin embargo, que muchos de los hablantes quedan muy alejados de la curva que indica la ecuación, como lo demuestra la poca fortaleza de la relación ( $R^2 = 0.085$ , sig. = 0.046).

Lo que ocurre es que las diferencias de edad son significativas solo entre los hablantes universitarios: aquellos (especialmente las mujeres) nacidos entre los cincuenta y los setenta del siglo pasado alcanzan puntuaciones más altas que los nacidos antes o después de esa época (Gráfico 6). La correlación entre las variables es cúbica ( $F = 6.508$ , sig. 0.003,  $R^2 = 49\%$ ) y explica un porcentaje de la

variable léxica muy superior al explicado por la correlación lineal ( $F = 1.259$ , sig. 0.274,  $R^2 = 5\%$ ).

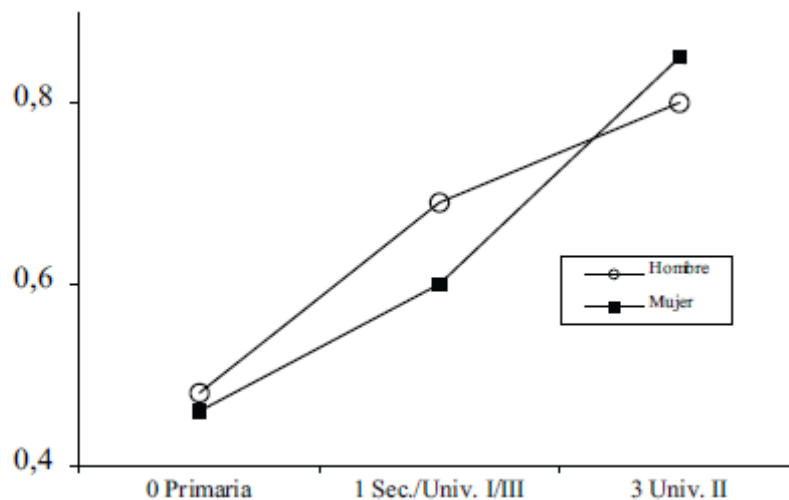
**Gráfico 6. Comparación del ajuste lineal y cuadrático entre la descentralización léxica y la edad entre los hablantes del grupo educacional universitario**



### 3.2.3 INTERACCIÓN

Los patrones cruzados de edad, sexo y educación recogen con simplicidad las regularidades observadas y producen unos resultados mucho más claros y tajantes: la descentralización léxica asciende conforme asciende el nivel educacional de los hablantes de ambos sexos; sin embargo, este ascenso sufre una inflexión entre los universitarios del grupo de edad intermedio (II), especialmente entre las mujeres (Gráfico 7).

Gráfico 7. Diferencias de sexo, edad y educación en la descentralización léxica



### 3.3 ANÁLISIS MULTIVARIANTE

Los modelos de variación que incluyen conjuntamente todas las variables de hablante utilizadas hasta aquí permiten saber cuáles de ellas tienen efecto significativo en la determinación de la variación léxica, así como si su efecto es redundante y, por tanto, prescindible en el modelo. Se realizaron tres análisis diferentes: de regresión lineal (3.3.1), multicausal (3.3.2) y de componentes principales (este último no se incluye aquí por razones de espacio). El primero permite seleccionar las variables con efectos independientes y significativos sobre la variable léxica (Etxebarria 1999; Hernández Campoy y Almeida 2005) el segundo reconstruye hipotéticamente el escenario sociorreticular de la variación lingüística mediante la consideración de los efectos directos e indirectos de las variables predictoras sobre la variable léxica (cf. Bisquerra; García Ferrando 1985; Martínez Arias 1999, Oakes 1998).

#### 3.3.1 EDUCACIÓN E INTENSIDAD RETICULAR

Al considerar conjuntamente el efecto de todas las variables de hablante, solo la educación y la intensidad de los vínculos reticulares demostraron tener un

efecto significativo en la predicción de la descentralización léxica (Tabla 5). El resto de las variables resultaron excluidas del modelo debido a su poca fortaleza o a su redundancia con respecto a las variables seleccionadas.

**Tabla 5. Efecto de las variables de hablante sobre el ICoDD. Análisis de regresión múltiple por pasos**

Términos en la ecuación de regresión	Modelo (1)	Modelo (2)
Correlación múltiple	0.639	0.703
R <sup>2</sup>	40.8%	49.5%
Constante	0.391	0.476
Educación	0.639 (6.644)	0.675 (7.479)
Intensidad		-0.297 (3.288)

Las probabilidades para los efectos están basadas en una “prueba de una cola” (one-tail test) a un nivel de confianza de 0.05. Se indican entre paréntesis los resultados de los t-tests para la regresión de los cuadrados mínimos ordinarios. En todos los casos la sig. = 0.000, salvo en el de Intensidad (sig. = 0.002)

Los efectos combinados de la educación y la intensidad reticular explican el 49.5% de la variación léxica estudiada. La probabilidad de descentralización aumenta a medida que aumentan los años de estudio cursados por el hablante (en la proporción de 0.639 unidades tipificadas de ICoDD por cada unidad tipificada de estudios). Cuando se tiene en cuenta la fuerza de los vínculos (y por tanto la integración del hablante en su red personal), esta influye negativamente (-0.297) y, además, eleva el efecto de la educación (0.675): cuanto más fuertes son los vínculos del hablante con su comunidad local menos descentralizado su vocabulario y mayor la importancia de los estudios.

### 3.3.2 EDUCACIÓN Y COMPETENCIA LÉXICA ESPECIALIZADA

Para representar adecuadamente la hipótesis de un efecto multicausal de las variables predictoras sobre ICoDD en la que se tenga en cuenta el control de unas variables sobre otras, se construyeron dos modelos causales (*path analysis*) basados en coeficientes tipificados Beta (*path coefficients*) de las correspondientes

ecuaciones de regresión (ecuaciones estructurales), con el objeto de medir los efectos directos e indirectos de aquellas sobre la descentralización léxica.

El primer modelo responde a la hipótesis sociológica clásica del efecto directo de la posición social sobre la capacidad lingüística. Se construyeron tres ecuaciones con tres variables endógenas (esto es, variables influidas o controladas por otras): la educación, la ocupación y el ICoDD. El modelo explica el 56.0% de la variación y las variables más influyentes en los efectos totales fueron la educación, la educación de la madre del hablante y la intensidad de los vínculos de su red personal.

El segundo modelo incluye la capacidad léxica especializada (Pident) y el ICoDD como variables endógenas (Gráfico 8). La hipótesis ahora es que el efecto de las variables de hablante (como la ocupación, la educación o la intensidad reticular) no es directo sobre el ICoDD, sino indirecto a través de la competencia especializada (Pident), que se erige así como la variable más determinante en el sentido de que no es la posición social la que influye en la mayor o menor competencia léxica del hablante, sino si esa formación se ha enfocado con posterioridad hacia la adquisición y comprensión de vocabulario poco corriente (literario, especializado, técnico, etc.).

El efecto total más importante en el segundo modelo (Tabla 6) es el de la educación del hablante (0.71) y de Pident (0.68), pero se trata, en el primer caso, del efecto de los años de estudio controlados por el grado de conocimiento especializado del hablante (Pident). Al margen de este, el efecto de la educación sobre ICoDD es pequeño (0.15). Las demás variables apenas tienen importancia.

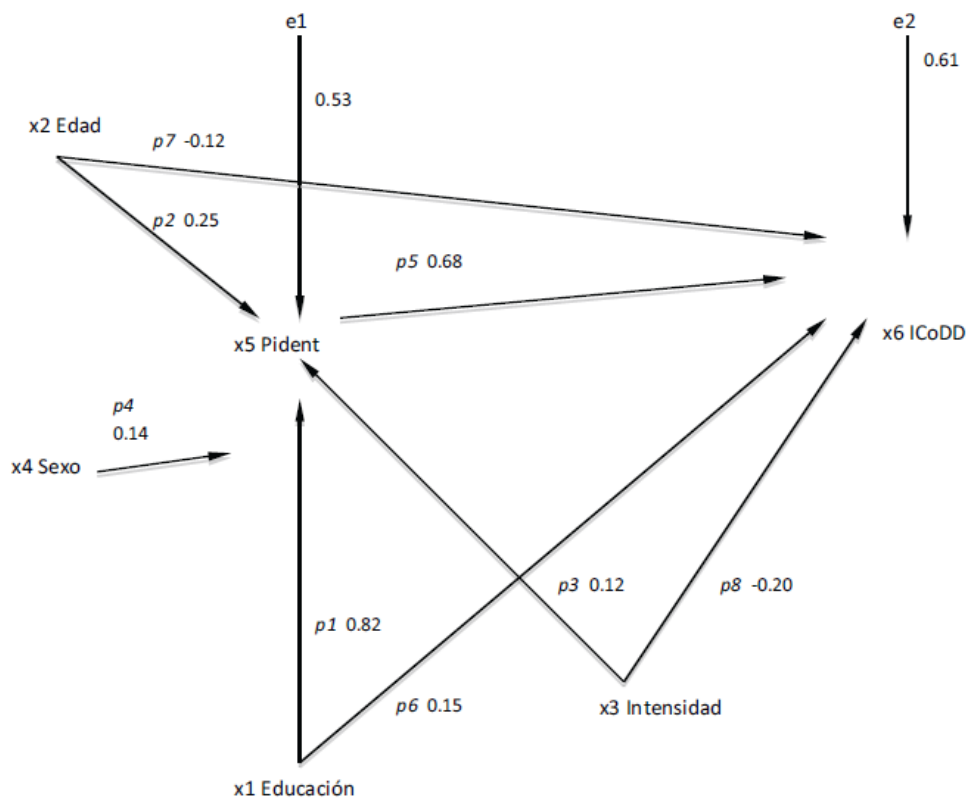
**Tabla 6. Efectos directos, indirectos y totales de las variables de hablante sobre**

ICoDD			
	Total	Indirecto(Pident)	Directo
Efecto			
x1Educación	0.71	0.56	0.15
x2Edad	0.05	0.17	-0.12
x3Intensidad	-0.12	0.08	-0.20
x4Sexo	0.10	0.10	—
x5Pident	0.68	—	0.68

Se ha recodificado el sexo del hablante como una variable ficticia (*dummy variable*) en la que H = 0 y M = 1

Este segundo modelo, que integra una variable de intervención léxica, explica el 63.0% de la varianza de ICoDD, frente al mencionado 56% del primer modelo (cuyas dos variables de intervención son la educación y el prestigio ocupacional) y, por tanto, es más simple y con mayor capacidad explicativa.

**Gráfico 8. Diagrama del proceso causal de los efectos lingüísticos y extralingüísticos sobre ICoDD**



VARIABLES

Exógenas:  $x1$  Educación;  $x2$  Edad;  $x3$  Intensidad;  $x4$  Sexo

Endógenas:  $x5$  Pident;  $x6$  ICoDD

Ecuaciones estructurales

$$(1) \text{ Pident} = x1 \text{ Educación} + x2 \text{ Edad} + x3 \text{ Intensidad} + x4 \text{ Sexo} + e1$$

$$(2) \text{ ICoDD} = x5 \text{ Pident} + x1 \text{ Educación} + x2 \text{ Edad} + x3 \text{ Intensidad} + e2$$

Error =  $1 - R^2$

$$(1) \text{ Pident } e1 = 0.53$$

$$(2) \text{ ICoDD } e2 = 0.61$$

$$\text{Varianza explicada} = 1 - (e2)^2 = 0.625$$

## 4 CONCLUSIONES

### 4.1 *COMPETENCIA LÉXICA*

Disponemos ahora de un modelo coherente de la capacidad léxica individual obtenido a partir de los datos suministrados por las pruebas de disponibilidad. Este se basa en la descentralización del léxico aportado por los hablantes.

### 4.2 *CONDICIONAMIENTO SOCIAL*

La descentralización léxica individual se explica en buena medida teniendo en cuenta los años de estudio cursados por los hablantes (que determinan casi completamente el prestigio social de sus ocupaciones). Sin embargo, el efecto de la educación sobre la capacidad léxica es pequeño cuando se considera el efecto de la capacidad léxica especializada; esto es, la capacidad de comprensión del vocabulario poco corriente (técnico, especializado, literario, etc.) adquirida con posterioridad al período de formación académica. Esta última capacidad se correlaciona positivamente con la edad y tiene una notable tendencia femenina.



## BIBLIOGRAFÍA CITADA

- ÁVILA MUÑOZ, A. M., 2006. *Léxico disponible de los estudiantes preuniversitarios malagueños*, Málaga: Universidad.
- y J. M. SÁNCHEZ SÁEZ, 2010. «La disponibilidad léxica. Antecedentes y fundamentos», en Ávila y Villena (eds.) (2010).
- y J. A. VILLENA PONSODA (eds.), 2010. *Variación social del léxico disponible en la ciudad de Málaga. Diccionario y análisis*. Málaga: Sarriá.
- y J. M. SÁNCHEZ SÁEZ, 2011. «La posición de los vocablos en el cálculo del índice de disponibilidad léxica: procesos de reentrada en las listas del léxico disponible de la ciudad de Málaga», *ELUA* 25: 45-74
- BISQUERRA ALSINA, R., 1989. *Introducción conceptual al análisis multivariable. Un enfoque informático con los paquetes SPSS-X, BMDP, LISREL y SPAD (I y II)*. Barcelona: PPU.
- ETXEBARRIA, J., 1999. *Regresión múltiple*. Madrid: La Muralla.
- GARCÍA FERRANDO, M., 1985. *Socioestadística. Introducción a la estadística en sociología*. Madrid: Alianza Universidad.
- BUTRÓN, G., 1987. *El léxico disponible: índices de disponibilidad*. Río Piedras: Universidad de Puerto Rico (Tesis Doctoral inédita).
- GOUGENHEIM, G., R. MICHEA, P. RIVENC Y A. SAUVAGEOT, 1956. *L'élaboration du français élémentaire. Étude sur l'établissement d'un vocabulaire et d'une grammaire de base*. París: Didier.
- , 1964. *Français Fondamental (1er degré). Étude sur l'établissement d'un vocabulaire et d'une grammaire de base. Nouvelle édition refondue et augmentée*. París: Didier.
- LÓPEZ CHÁVEZ, J. Y C. STRASSBURGER FRÍAS, 1987. «Otro cálculo del índice de disponibilidad léxica», en *Presente, y perspectiva de la investigación computacional en México. Actas del IV Simposio de la Asociación Mexicana de Lingüística Aplicada*. México: UNAM.
- LÓPEZ MORALES, H., 1999. *Léxico disponible de Puerto Rico*. Madrid: Arco/Libros.
- LORÁN, R., 1983. *Un índice de disponibilidad léxica*. Mayagüez: Departamento de Matemáticas. Universidad de Puerto Rico.
- y H. LÓPEZ MORALES, 1983. *Nouveau calcul de l'indice de disponibilité*. Puerto Rico: ms.
- MARTÍNEZ ARIAS, R., 1999. *El análisis multivariante en la investigación científica*. Madrid: La Muralla.
- OAKES, M., 1998. *Statistics for corpus linguistics*. Edimburgo: Edinburgh University Press.
- VILLENA PONSODA, J. A. y A. M. ÁVILA MUÑOZ, 2010. «Condicionantes sociales del léxico disponible», en Ávila y Villena (eds.) (2010).
- VILLENA PONSODA, J. A., A. M. ÁVILA MUÑOZ y J. M. SÁNCHEZ SÁEZ, 2012. «Condicionamiento social de la capacidad léxica individual. Patrones de sociolingüísticos del vocabulario disponible en la ciudad de Málaga», en Villena y Ávila (eds.) (2012).
- VILLENA PONSODA, J. A. y A. M. ÁVILA MUÑOZ, 2012: *Estudios sobre el español de Málaga: pronunciación, vocabulario, sintaxis*. Málaga: Sarriá.
- ZADEH, L. A., 1965. «Fuzzy sets», *Information and Control* 8: 338-353.
- ZIMMERMANN, H. J., 2001. *Fuzzy Set Theory and its Applications*. Boston: Kluwer Academic Publishers.