

EL LEXICO DISPONIBLE: INDICES DE DISPONIBILIDAD

Disertación presentada a la Facultad del Departamento de  
Estudios Hispánicos, como uno de los requisitos para  
obtener el grado de Doctor en Filosofía y Letras,  
en la Universidad de Puerto Rico

Diciembre de 1987

Glória Izáskun Butrón

Aprobada con la calificación de *Sobresaliente por unanimidad*

Presidente del Comité Examinador

## INTRODUCCION

El léxico de una lengua puede estudiarse desde diversas perspectivas pero, desde todas ellas las herramientas de tipo cuantitativo han resultado de especial utilidad.

La léxico-estadística tiene en su haber numerosos estudios en los cuales se realizan ordenamientos léxicos basados en la frecuencia de las palabras. A partir de esos estudios sabemos que las palabras más frecuentes -aquéllas que se repiten en cualquier discurso y ocasión- pertenecen a inventarios cerrados, son las llamadas palabras gramaticales, como artículos, preposiciones, pronombres, etc.. Estas son consideradas los términos más estables. Sin embargo, y también gracias a los mismos estudios nos percatamos de que hay palabras por todos conocidas, que designan la realidad inmediata del hablante, y que no alcanzan frecuencias relevantes. Por tal razón se propone otro tipo de estudio con el objeto de determinar esas palabras. La innovación consistió en emplear la frecuencia léxica pero, en relación con campos referenciales específicos, que se denominaron "centros de interés".

Los estudios léxicos bajo esta perspectiva de campos específicos se denominan de "disponibilidad léxica" y son mucho menos numerosos que los conocidos diccionarios de frecuencias. El concepto de disponibilidad léxica fue elaborado en la década del 50 y se pone en práctica por

primera vez para un estudio del francés<sup>1</sup>. Durante bastante tiempo, aunque se cambia la categoría de las palabras analizadas, ya que los investigadores se interesan por aquéllas menos estables (los nombres), la única forma de reflejar la disponibilidad que las palabras tienen para los informantes, continuó siendo la frecuencia de menciones.

En 1983 se propone un nuevo método que cuantifica la disponibilidad y, que no sólo considera la frecuencia, sino que toma en cuenta, además, el orden en que las palabras son mencionadas por el informante. Este método consiste en el cálculo de un "índice de disponibilidad" mediante una fórmula matemática que combina la frecuencia y el orden de mención. Fue propuesto por López Morales y Lorán<sup>2</sup> y se utiliza en diversos trabajos de investigación<sup>3</sup>, realizados en diversos países; los comentamos en el recuento bibliográfico que incluimos en el Capítulo I.

En 1986 un grupo de investigadores de la Universidad

---

1. G. Gougenhém et al., L'Elaboration du français fondamental (1er. degré), Didier, París, 1964.

2. H. López Morales y R. Lorán, "Nouveau calcul de l'indice de disponibilité", en prensa, 1983.

3. B. Román, Disponibilidad léxica en escolares de Dorado, (tesina inédita), Río Piedras, Universidad de Puerto Rico, 1985.

H. G. Justo Hernández, Disponibilidad léxica en colores, (tesina inédita), México, Universidad Nacional Autónoma de México, 1986.

M. Mena Osorio, Disponibilidad léxica infantil en tres niveles de enseñanza básica, (tesis inédita), Concepción, Chile, Universidad de Concepción, 1986.

A. V. Canizal Arévalo, Disponibilidad léxica en escolares de primaria terminada. Análisis de seis centro de interés, (tesina inédita), México, Universidad Nacional de México, 1987.

Nacional Autónoma de México al trabajar con la fórmula propuesta por López Morales y Lorán encuentra ciertas irregularidades en los resultados, ante lo cual proponen una nueva fórmula para calcular el índice de disponibilidad<sup>4</sup>.

Nuestra intención en este trabajo es analizar desde dos puntos de vista -el teórico y el empírico- las diferentes fórmulas. Para el enfoque empírico empleamos una muestra de estudiantes de escuela elemental recogida por nosotros. Para el diseño de la muestra consideramos variables sociolingüísticas que nos permiten llevar a cabo comparaciones entre los resultados arrojados por las diferentes fórmulas.

Hasta el momento, no se ha realizado ningún estudio comparativo entre los órdenes que generan las diferentes fórmulas, por lo cual a esta tarea nos avocamos por medio de la fórmula de Spearman, frecuentemente usada en estadística para medir la correlación entre ordenaciones o rangos.

En nuestro trabajo también proponemos nuevas aplicaciones de los conceptos de "disponibilidad" y de "índice de disponibilidad". Definimos nuevos índices para un informante y para un grupo de informantes. Estas aplicaciones aún no han sido exploradas por ningún investigador, hasta el momento de la redacción de estas líneas. Con excepción de una fórmula brevemente esbozada

---

4. J. López Chávez y C. Strassburger Frías, Otro cálculo del índice de disponibilidad léxica, México, Universidad Nacional Autónoma de México, 1987.

por López Chávez y Strassburger<sup>5</sup> para calcular el "índice de disponibilidad de un informante" no conocemos ningún otro trabajo en esa dirección.

Aplicamos además a nuestra muestra una medida de dispersión, propuesta por M. S. Echeverría<sup>6</sup>, que pretende reflejar la cohesión en centros de interés, en algunos de los centros de interés escogidos por nosotros.

En resumen, pretendemos con este estudio realizar un avance en el empleo de herramientas cuantitativas en los estudios léxicos, clarificando situaciones ya planteadas, presentando nuevas interrogantes y proponiendo algunas soluciones.

El uso creciente de las herramientas y técnicas que aquí discutimos es una prueba más de la importancia que éstas tienen en la investigación y planificación lingüística, en particular en la planificación curricular que tanto nos preocupa, así como en estudios dialectales, sociolingüísticos y psicolingüísticos.

---

5. Ibid., pag. 23.

6. M. S. Echeverría et al., Disponibilidad léxica en Educación Media, Concepción, Chile, Universidad de Concepción, 1987.

## CAPITULO I

### ESTUDIOS CUANTITATIVOS

"Todo está en la palabra... Una idea entera se cambia porque una palabra se trasladó de sitio, o porque se sentó como una reinita adentro de una frase que no la esperaba y que la obedeció... Tienen sombra, transparencia, peso, plumas, pelos, tienen de todo lo que les fue agregado de tanto rodar por el río, de tanto transmigrar de patria, de tanto ser raíces..."

Pablo Neruda

1. Generalidades. La lengua española consta de un vocabulario inmenso y se nutre constantemente de aportes científicos y técnicos en la compleja civilización en que vivimos. Surgen, pues, divergencias entre los diccionarios y la norma léxica de una determinada comunidad de habla. Desaparecen de la cultura viva los objetos o ideas que algunas palabras designaban antaño y que sin embargo siguen presentes en las hojas de los diccionarios. Hay desgaste de palabras tradicionales que al volverse inexpresivas para el hablante las sustituye por neologismos. Constantemente se crean regionalismos, palabras que sólo se conocen en alguna zona o país del mundo hispánico.

Al ser el léxico lo más externo y cambiante de la lengua ha sido objeto desde el siglo XV de numerosos estudios. Ese interés por el estudio del vocabulario se ha visto incrementado con el tiempo, ya que los cambios léxicos se suceden y son cada vez más bruscos debido a los modernos sistemas de comunicación.

Resulta imprescindible saber cuál es realmente el vocabulario que usamos y cuál aquél que permanece dormido en los diccionarios. Y aún más en qué orden debe integrarse ese vocabulario al lenguaje infantil.

Ante estas interrogantes es que nos planteamos iniciar nuestra investigación. Consideramos que hacen falta métodos fáciles y accesibles para el estudio del vocabulario vivo de nuestra lengua. Pensamos que para mantenernos al día con estos cambios cuando estamos considerando una lengua que es hablada por trescientos millones de personas en muchos países, puede sernos de gran ayuda utilizar las nuevas tecnologías, puestas a nuestro alcance con la informática, que facilitarán enormemente la tarea.

## 2. Antecedentes en el tema: la léxico-estadística.

Desde hace bastante tiempo y debido a su naturaleza contable, se vienen realizando estudios cuantitativos de la lengua. El manejo y clasificación de datos lingüísticos se ha hecho utilizando criterios dependientes de razones, percentilas, medidas de centralización, medidas de dispersión y otros métodos estadísticos. Mediante ellos ha sido posible un empleo intensivo de los instrumentos estadísticos y los computadores en beneficio de los estudios léxicos en el marco de la llamada léxico-estadística.

Los primeros recuentos léxicos aparecen ante la importancia de conocer el vocabulario que manejan los hablantes de una lengua determinada. La mayor parte de

estos recuentos tienen una finalidad pedagógica y muy frecuentemente planteamientos teóricos muy débiles. Es importante señalar que la mayor parte de estos recuentos tampoco coinciden en la segmentación de las unidades que van a ser medidas.

Si bien no hay dudas de que la unidad básica de la léxico-estadística es la palabra, durante mucho tiempo y hasta el presente este término ha tenido muchas interpretaciones. Entre los planteamientos teóricos más serios son de mencionar los presentados en el recientemente publicado Léxico básico del español de Puerto Rico<sup>1</sup>.

Para la escuela estructuralista, por ejemplo, el término palabra resulta de tal ambigüedad que prefiere utilizar el morfema como elemento de análisis. Sin embargo, la léxico-estadística debe medir una unidad mínima independiente que es también la unidad que el hablante común y corriente puede identificar.

Algunos estudiosos toman cada forma flexiva como unidad independiente, otros prefieren la "unidad léxica", incluyendo las flexiones con cálculos separados o no. Los fenómenos que deben considerarse son la polisemia y homografía, división y agrupación de cadenas y asignación de palabras a clases. Son las perífrasis, por ejemplo, las que más discusiones y desacuerdos han ocasionado. Aunque como ya hemos señalado, pocos recuentos fijan de antemano sus

---

1. A. Morales, Léxico básico del español de Puerto Rico, San Juan, Academia Puertorriqueña de la Lengua Española, 1986.



criterios de segmentación.

Coincidimos con Amparo Morales<sup>2</sup> en que la palabra puede interpretarse atendiendo a tres niveles diferentes: fonológico, gramatical y semántico; y bajo la perspectiva de dos planos diferentes: sintagmático y paradigmático.

A nivel fonológico se caracteriza por la forma de sus unidades fónicas a nivel segmental. En el gramatical, se atiende a la forma y función en la estructura oracional. Semánticamente debe considerarse la palabra como unión de una forma y un contenido.

En el plano sintagmático las características principales serán la coherencia interna y la independencia distribucional. A nivel de paradigma se define por unas cualidades que la identifican por su forma y contenido y oponen su funcionamiento al del resto de las unidades del sistema.

En léxico-estadística, se suele emplear el concepto de palabra desde una perspectiva formal, como el material gráfico que se encuentra limitado por espacios en blanco. Esta unidad se puede identificar, aislar y contar fácilmente, y favorece la explicación de las propiedades distribucionales de sus constituyentes.

Ya a partir del estudio estadístico de Kaeding<sup>3</sup> en 1897

---

2. A. Morales, Teorías y métodos aplicables a la Estadística-lingüística del español americano, tesis doctoral inédita, Universidad de Puerto Rico, 1975.

3. J. W. Kaeding, Häufigkeitwörterbuch der deutschen sprache, Steglitz bei Berlin, der Herausgeber, 1897.

se vio cómo unas pocas palabras -las más frecuentes-<sup>4</sup> representaban la mayor parte del texto. Los diccionarios estadísticos incluyen en general listas de palabras o lemas ordenados por su frecuencia de mención en un texto o grupo de éstos. Listas lematizadas son aquéllas en que se toma como unidad la unidad léxica y las formas flexivas se consideran parte de la misma en los cómputos generales.

En estos trabajos se usa fundamentalmente el criterio de listas de frecuencia, es decir, palabras ordenadas según el número de apariciones en textos determinados. Varían los criterios de selección de los textos (todos ellos escritos), y también el tamaño de las muestras.

Entre los diccionarios de frecuencias, están tanto los sincrónicos como los diacrónicos. El primero de ellos parece ser el mencionado de Kaeding para el alemán. En francés suele citarse como primero al de Henmon<sup>5</sup>, publicado en 1924. Con respecto al español el primero fue el de García Hoz<sup>6</sup> de 1952 y en el mismo año apareció el trabajo pionero de Rodríguez Bou<sup>7</sup>, que recoge el recuento de

---

4. Las 15 palabras más frecuentes representaban 25% del total, las 66 palabras más frecuentes representaban casi el 50%.

5. V. A. C. Henmon, A French Work Book Based on a Count of 400,000 running words, Madison, Bureau of Educational Research, University of Wisconsin, 1924.

6. V. García Hoz, Vocabulario usual, vocabulario común y vocabulario fundamental, Madrid, CSIC, 1952.

7. I. Rodríguez Bou, Recuento del vocabulario español, I y II, Río Piedras, Editorial Universitaria, 1952.

Buchanan<sup>8</sup> de 1916. Otros trabajos para el español fueron los de Keninston<sup>9</sup>, Céspedes<sup>10</sup>, Jameison<sup>11</sup> y Catwright<sup>12</sup>. Generalmente estos trabajos están basados en una enorme cantidad de datos, que hacen de su elaboración un proceso largo y costoso.

Estos son trabajos muy valiosos para la época. No es fácil compararlos entre sí puesto que siguen criterios muy discímiles. Todos ellos tienen como marco de referencia para la muestra la literatura en un sentido bien amplio; pero no toman en cuenta las diferencias de época de las obras, mezclando obras actuales y documentos de épocas lejanas. También difieren en cuanto a la selección de las mismas. Emplean desde obras literarias reconocidas hasta escritos escolares. Este es el caso del Recuento del vocabulario español de Rodríguez Bou publicado por el Consejo Superior de Enseñanza. Al seleccionar las obras tampoco se tiene en cuenta el origen de sus creadores; por ejemplo, si han

---

8. M. Buchanan, A graded spanish word book, Toronto, Toronto University Press, 1941.

9. H. Keniston, Basic List of Spanish words and idioms, Chicago, The University of Chicago Press, 1928.  
H. Keniston, Spanish Idiom List, selected on the basis of frequency of occurrence, New York, Macmillan Co., 1929.

10. T. R. Céspedes, Investigación acerca de las palabras usadas en castellano, Panamá, Star and Herald, 1928-29.

11. E. Jameison, "A standardized vocabulary for elementary spanish", MLJ, 8, 1924.

12. C. W. Cartwright, "A study of the vocabularies of eleven Spanish grammars and fifteen Spanish reading texts", MLJ, 10, 1925.

sido escritas por españoles o puertorriqueños olvidando que es muy difícil llegar a conclusiones acerca del uso de una lengua sin hablar de usuarios determinados. Podemos hablar de una lengua común -el español- pero con características léxicas muy diversas. Lo que se trata de identificar son las palabras más usadas por unos hablantes determinados que basan sus diferencias en características geográficas, culturales, sociales, etc..

La tónica que prevalece en los vocabularios de frecuencia en español en cuanto a la técnica de muestreo usada es una total ausencia de postulados a priori respecto a la población que consideran objeto de su estudio, los criterios de representatividad de las muestras son subjetivos y las categorías o subgrupos corresponden a un proceso posterior a la selección<sup>13</sup>.

La preparación de léxicos básicos ha recibido una atención creciente por parte de pedagogos y lingüistas. El criterio que primeramente se eligió para el establecimiento de estos léxicos fue también el de la frecuencia.

Un gran paso adelante en la léxico-estadística fue integrar la idea de dispersión al concepto de frecuencia como fuera hecho por A. Juilland en su colección de léxicos básicos de las lenguas románicas<sup>14</sup>.

---

13. A. Morales, Op. cit.

14. A. Juilland y E. Chang Rodríguez, Frequency Dictionary of Spanish Words, The Romance Languages and their Structures, First Series SI, The Hague, Mouton, 1964.  
A. Juilland y P. M. A. Edwards e I. Juilland, Frequency Dictionary of Rumanian Words, The Hague, Mouton, 1965.

La frecuencia de las palabras está muy influida por los temas que se desarrollan en los textos seleccionados. Si, por ejemplo, escogemos un texto relacionado con la medicina pueden aparecer en el mismo infinidad de palabras que un hablante corriente no utilizaría jamás en toda su existencia como podría ser "pregangliónicas". Así, en lugar de agrupar todo tipo de textos y sacar frecuencias generales de las palabras que lo integran, se divide el universo léxico en diferentes mundos, delimitados, ya sea por el contenido de los textos o las condiciones formales de los mismos, y se observa si una palabra tiene la misma o diferente frecuencia en cada uno de éstos. A esto se le denomina dispersión.

La lengua escrita es la utilizada en estos estudios pues ofrece más posibilidades de rigurosidad, ya sea en cuanto al tipo de texto, como a la exhaustividad y sincronía que se requiere para elegir la muestra. En general se coincide en dividir el universo léxico en cinco mundos: drama, narrativa, ensayo, literatura técnica y periodismo.

Con el propósito de neutralizar el peligro de no tener en cuenta la distribución de las frecuencias en los textos se emplea el concepto de rango o dispersión según la

---

A. Juilland, D. Brodin y C. Davidovitch, Frequency Dictionary of French Words, The Hague, Mouton, 1971.

A. Juilland y V. Traversa, Frequency Dictionary of Italian Words, The Hague, Mouton, 1973.

terminología empleada por Juilland<sup>15</sup>. Mediante cálculos matemáticos se combina la frecuencia con la dispersión y se obtienen así unos índices llamados de dispersión compleja y de uso.

Recientemente se ha publicado el Léxico Básico del Español de Puerto Rico<sup>16</sup>; en él se implementa la misma fórmula empleada por Juilland en su léxico básico del italiano, con algunos retoques que mejora la empleada en el Frequency Dictionary of Spanish Words.

Ningún recuento de los principales trabajos en léxico-estadística podría dejar de mencionar algunos trabajos de orientación más teórica y matemática, en particular, los estudios sobre los comportamientos de las frecuencias relativas, los cuales suponen unos índices de probabilidad o distribuciones de probabilidad regulares. La relación de Zipf<sup>17</sup>, establece el carácter constante del producto de la frecuencia de una palabra, por su rango elevado a una potencia alfa (constante para un corpus dado).

---

15. Para una explicación pormenorizada de las fórmulas de frecuencia, dispersión y uso utilizadas en los diccionarios más importantes del español, véase A. Juilland y E. Chang-Rodríguez, "Frequency Dictionary of Spanish Words", The Romance Languages and their Structures, First Series SI, The Hague, Mouton, 1964. p. XXXVIII y ss.. También véase V. Bortolino, C. Tagliavini y A. Zampolli, Lessico di frequenza della Lingua Italiana Contemporanea, Italia, IBM, 1972. p. XXIII y ss..

16. A. Morales, Op. Cit.

17. Para información más detallada de la relación de Zipf véase el estudio de M. Moreau, Bulletin d'Information du Laboratoire d'Analyse Lexicologique, 6, Centre d'Etude du Vocabulaire Français de la Faculté des Lettres et Sciences Humaines de Besançon, 1961.

La ley de Simon de 1955, establece los postulados de que la probabilidad de que una palabra sea la siguiente en aparecer es directamente proporcional a su frecuencia de aparición anterior; y que la probabilidad de que la palabra siguiente sea una palabra nueva se mantiene constante.

Otra relación que debe mencionarse es la ley de Mandelbrot que data de 1953; ésta modifica la de Zipf estableciendo una relación curvilínea entre la frecuencia y el rango de una palabra en el extremo de los rangos superiores.

El modelo de dos grupos de Mandelbrot es también modificado por la ley de Herdan de 1964, que establece un modelo de tres grupos de distribución: el de los rangos inferiores, el de los medios y el de los superiores, siendo el de los rangos medios el más variable entre muestras de distinto tamaño.

Pero todavía no se ha resuelto este problema de la distribución de la frecuencia, porque aún no se ha logrado el modelo que establezca un índice o relación, o función de distribución general (es decir, independiente de las distintas zonas de rango), y porque todavía no se ha establecido la función de distribución con independencia del tamaño de la muestra. De todos modos, se ha buscado la comparabilidad entre los resultados de los distintos trabajos sobre frecuencias del léxico, realizados con distintos tamaños entre sí, a través de la introducción del tratamiento no puntual sino por intervalos de la escala de

frecuencia, de modo que las posiciones relativas de las palabras en la escala a intervalos se mantenga constante. Las frecuencias absolutas se transforman en valores logarítmicos.

3. Más antecedentes: la disponibilidad léxica. Otro criterio que se utiliza crecientemente en los estudios cuantitativos del léxico, es el de la disponibilidad léxica. Está basado en hechos tanto lingüísticos como sicolingüísticos.

Fue Michéa el primer lingüista que hizo encuestas de disponibilidad léxica. El la describe así<sup>18</sup>:

"En présence d'une situation donnée, les mots que viennent le premiers à l'esprit sont ceux qui sont liés tout spécialement à cette situation et la caractérisent". "Un mot disponible est un mot qui, sans être particulièrement fréquent, est cependant toujours prêt à être employé, et se présente immédiatement à l'esprit o ù momant ou l'on en a besoin."

La teoría de la disponibilidad léxica está basada en las teorías asociacionistas de la sicolingüística que estudian las relaciones entre palabras, tanto entre sí, como con otros conceptos e imágenes asociados significativamente.

Este comportamiento psicológico subyacente al comportamiento lingüístico, viene refrendado por ciertos hechos comprobados como, por ejemplo: los adultos tienden más a las asociaciones paradigmáticas (respondiendo a la

---

18. R. Michéa, "Mots fréquents et mots disponibles. Un aspect nouveau de la statistique du langage", Les Langues Modernes, 47, 1953, p. 342.



palabra estímulo con otras palabras de la misma clase gramatical); los niños tienden más a las asociaciones sintagmáticas (es decir, a palabras que podrían ser consecutivas a la palabra-estímulo, y por eso pertenecer a otra clase gramatical).

Hay toda una tipología de las asociaciones, desarrollada por los sicolingüistas, y que en parte ha sido utilizada también por los lingüistas. Concretamente es la clasificación de "asociaciones libres/ asociaciones controladas". Los procedimientos de las asociaciones controladas son los que más se han aplicado a los trabajos de disponibilidad léxica, y que nosotros también empleamos en nuestra investigación.

Al examinar las listas de frecuencia, por ejemplo, en el Francés Fundamental<sup>19</sup> se encontró que había palabras muy comunes con una frecuencia bajísima y más extraño aún, palabras que todo el mundo conocía como lettre o timbre no aparecían en las listas.

De allí surge el concepto de disponibilidad. Es decir hay palabras que tienen una probabilidad bajísima o nula de aparecer, a menos que se den determinados temas. Por tanto, la disponibilidad léxica se entiende como el caudal léxico utilizable en una situación comunicativa dada<sup>20</sup>.

Son, pues, distintos los conceptos de léxico frecuente,

---

19. G. Gougenhém et al., Op. Cit..

20. H. López Morales, Enseñanza de la lengua materna, Madrid, Playor, 1984, p. 62.

situación frecuente, léxico disponible, situación determinada y léxico fundamental. Son palabras frecuentes aquéllas que independientemente del tema aparecen continuamente. Son palabras disponibles aquéllas que únicamente aparecen en determinadas situaciones. Se da la situación de que en las distintas categorías léxicas tenemos como palabras más frecuentes, en principio, las llamadas palabras gramaticales, luego los verbos también son muy estables, los adjetivos y los adverbios nos sirven también para comunicar infinidad de situaciones, pero las menos estables son los nombres. Es fácil de deducir que la palabra ola se utilizará únicamente cuando nos estemos refiriendo a onda que se produce en el mar, mientras que grande podemos aplicarlo calificando a ola, niño, casa etc..

Los léxicos básicos recogen las palabras más frecuentes en tanto la disponibilidad recoge las palabras más inestables estadísticamente, pero también usuales.

El léxico básico y el léxico disponible conforman el léxico fundamental, o sea, el caudal léxico manejado por los hablantes en cualquier circunstancia. Ambos conceptos no son excluyentes sino que pueden ser vistos como complementarios.

Una forma de obtener palabras disponibles para un informante es proponerle situaciones concretas que Michéa<sup>21</sup>

---

21. Michéa plantea la noción, conocida y utilizada en pedagogía de centro de interés en su artículo en *Langues Modernes* y la recomienda como base para la enseñanza de nombres concretos. Gougenheim aplica la misma a un método de investigación en L'Elaboration...

denomina "centros de interés" y pedirle entonces que escriba palabras que se relacionen con dicha situación.

De un modo parecido a como se hizo en los estudios de frecuencia del léxico al recurrirse también al criterio de la dispersión, de modo que el índice de la dispersión funcionase a modo de factor corrector del índice de frecuencia, también en los estudios sobre la disponibilidad léxica se ha procurado introducir un índice de la "fuerza asociativa" de una palabra.

La "fuerza asociativa" se ve reflejada por el lugar que una palabra ocupa en la lista de palabras respuesta que ofrecen para cierto "centro de interés" los distintos informantes. Njock<sup>22</sup> utiliza como medida de la fuerza asociativa el "rango asociativo medio" de una palabra para los informantes encuestados, entendiendo por rango asociativo la posición que la palabra ocupa para cada informante.

La hipótesis sobre la que se basa la validez de este índice es que a mayor importancia o disponibilidad de una palabra para un individuo, más rápido la incluirá en su lista-respuesta del centro de interés.

En el Capítulo II nos referiremos nuevamente al rango

---

22. P. E. Njock, L'univers familial de l'enfant africain, Québec, CIRB, 1979. Tomado de M. J. Azurmendi Ayerbe, Elaboración de un Modelo para la descripción sociolingüística del bilingüismo y su aplicación parcial a la comarca de San Sebastián, I, San Sebastián, Caja de Ahorros Provincial de Guipúzcoa, 1983.

asociativo medio y lo emplearemos como punto de referencia para la comparación de diferentes índices de disponibilidad

La disponibilidad léxica ha abierto las puertas para la realización de diferentes tipos de estudios que Azurmendi<sup>23</sup> clasifica de la siguiente manera:

- A) Establecimiento de léxicos básicos.
- B) Trabajos de tipo más teórico.
- C) Para la Comparación Inter e Intralingüística.

En el primer tipo de trabajo -establecimiento de léxicos básicos- se aúnan varios criterios a un tiempo, pero se ha encontrado que la disponibilidad no puede ser dejada de lado puesto que refleja mejor el uso oral y es más representativa de las realidades sociales. Por otra parte implica análisis menos costosos, pues son necesarios corpora más pequeños que los que se hacen imprescindibles en los estudios que tienen en cuenta sólo la frecuencia.

Hay trabajos de tipo más teórico con relación a la disponibilidad. Es muy interesante destacar el estudio de Mackey<sup>24</sup>, que intenta establecer el tamaño de la muestra, y el tamaño de respuestas óptimos para los estudios de disponibilidad.

Entre los trabajos que buscan la comparación

---

23. M. J. Azurmendi Ayerbe, Elaboración de un modelo para la descripción sociolingüística del bilingüismo y su aplicación parcial en la comarca de San Sebastián, I, II, III, tesis doctoral publicada por Caja de Ahorros Provincial de Guipúzcoa, 1983.

24. W. F. Mackey, Le sondage dans les enquêtes de disponibilité, Québec, CIRB, 1971.

interlingüística o intralingüística es también de destacar el trabajo de Mackey<sup>25</sup> que se propone establecer la distancia intralingüística entre variedades geolingüísticas (el francés de Francia y el de Canadá) en la misma lengua. Este trabajo tiene fundamentalmente una finalidad pedagógica.

4. Nuestra investigación. El concepto de disponibilidad léxica es un paso importante en el estudio del léxico, pero a nuestro entender es necesario desarrollar un método de ordenar las palabras para cada "centro de interés" que no solamente tenga en cuenta la frecuencia de aparición en las listas-respuesta de los informantes sino que también refleje la posición que éstas ocupan en dichas listas. También sería conveniente que ese método de ordenar fuera lo suficientemente flexible para poder reflejar la importancia relativa de una mención en una posición respecto de una mención en una posición más avanzada.

La necesidad de ordenar las palabras obtenidas en encuestas escritas realizadas en torno a diferentes "centros de interés", teniendo en cuenta además de la frecuencia la jerarquización dada por los individuos, es mencionada por primera vez por H. López Morales<sup>26</sup>. El cataloga el problema como de interés fundamental para que no se pierda

---

25. W. F. Mackey et al., Le vocabulaire disponible du français, I, Paris, Didier, 1971.

26. H. López Morales y R. Lorán, Op. Cit.

información. Es un factor importante que debe ser tomado en consideración puesto que no hay duda de que son más disponibles aquellas palabras que acuden primero a la memoria y que, por tanto, aparecen en los primeros lugares en las listas ofrecidas por los informantes. A este aspecto se le denomina "grado de disponibilidad"; indica qué palabras son muy disponibles y cuáles menos, y nos sirve para ponderar matemáticamente la frecuencia. Al tomar en consideración además de la frecuencia el rango, tendremos más poder discriminatorio.

Desde un punto de vista intuitivo no cabe duda de que frente a una situación concreta aquellas palabras que primero acuden a nuestra memoria son más disponibles en relación con dicha situación que aquéllas que no hacen su aparición de inmediato. De existir una forma de medir la disponibilidad dicha medida debería ser mayor para las palabras que aparecieran primero que para las restantes.

A dicha medida de la disponibilidad se le ha venido denominando "índice de disponibilidad", aún antes de proponerse ninguna forma concreta de calcularlo.

La situación descrita se complica más cuando se piensa en resumir las medidas de la disponibilidad de los diferentes individuos en una medida única que represente a toda la población.

R. Lorán<sup>27</sup> propuso un método que permite ordenar el listado de palabras mediante el cálculo de un índice de

---

27. H. López Morales y R. Lorán, Ibid.

disponibilidad el cual será una representación del llamado orden de disponibilidad.

El presente trabajo se propone complementar el mencionado de R. Lorán, así como otros que han sido secuela de éste. Dicha complementación será no sólo a nivel teórico sino práctico. El propósito de nuestra investigación es partir del concepto de disponibilidad léxica, hacer una selección de "centros de interés" o estímulos adecuados, discutir los axiomas o metodologías que conducen a la fórmulas propuestas, compararlas y contrastar su validez.

La primera fórmula propuesta por Lorán tiene aplicación en muestras limitadas en el número de respuestas, es decir, en las que el tamaño de todas las listas-respuesta es igual para todos los informantes. Para obtener tales listas teníamos dos alternativas, bien en cada "centro de interés" cortar las listas según el informante que hubiera contestado menor número de palabras en el tiempo disponible (igual para todos) y despreciando las palabras en exceso de la mayoría de las listas, o bien solicitar a los informantes que respondieran todos un número predeterminado de palabras haciendo así las listas todas iguales. La primera alternativa fue la escogida luego de varias pruebas piloto.

La segunda fórmula es aplicable al caso de listas-respuesta de tamaño desigual. Para llevarla a la práctica dimos a los informantes un tiempo limitado en el cual debían escribir todas las palabras que fuesen capaces respecto de cada "centro de interés"; obtuvimos de esa manera listas de

tamaño diferente para cada informante.

Forma parte de ambas fórmulas un factor de ponderación de las posiciones denominado "tasa de sustitución". En nuestra investigación experimentamos dando diferentes valores a la tasa de sustitución y discutimos los resultados obtenidos a efectos de determinar cuál es el valor más idóneo.

En 1986 J. L. Chávez y C. Strassburger Frías proponen una simplificación de la fórmula anterior. Sin embargo en 1987 los mismos autores plantean que tanto en la fórmula de Lorán como en la modificada propuesta por ellos mismos la dependencia del número de informantes y la proporción que guardan en el descenso provoca ciertas incongruencias en los resultados y proponen una cuarta fórmula que, aunque no varía esencialmente el concepto de las originales, pretende resolver algunas incongruencias por ellos encontradas. Esta última fórmula sería válida tanto para muestras cerradas como abiertas<sup>28</sup>.

En nuestro trabajo aplicamos las tres fórmulas en situaciones reales, para llegar a algunas conclusiones tanto sobre su validez como sobre la metodología para su aplicación. Asimismo, comparamos los resultados obtenidos con un mismo corpus mediante la aplicación de cada una de

---

28. J. López Chávez y C. Strassburger Frías, Otro cálculo del índice de disponibilidad léxica, México, Universidad Nacional Autónoma de México, 1987.



ellas y la fórmula de Spearman<sup>29</sup>.

5. Organización del presente trabajo. Luego de una breve descripción del contenido de cada uno de los capítulos de este trabajo, concluimos este primero con una reseña en la que revisamos someramente la bibliografía disponible sobre el tema.

En el Capítulo II presentamos cuidadosamente la obtención de las diferentes fórmulas para el índice de disponibilidad, discutiendo en detalle las propiedades o axiomas que condujeron a ellas y efectuamos su comparación. Calculamos además la correlación entre las ordenaciones derivadas de las fórmulas LL y LS<sup>30</sup> así como de la frecuencia absoluta y el rango asociativo medio de Njock.

En el Capítulo III presentamos el índice de disponibilidad para un grupo de informantes y el índice de disponibilidad para una persona.

En el Capítulo IV analizamos la categoría léxica de los sustantivos en el lenguaje infantil de niños de quinto grado de escuela elemental, considerando distintas situaciones sociolingüísticas. Discutimos las distintas variables consideradas, así como el formato de la encuesta. Seguimos el criterio innovador de Michéa sobre disponibilidad léxica

---

29. C. I. Chase, Elementary statistical procedures, New York, Mc Graw Hill, 1984, p. 145.

30. Para simplificar al referirnos a los diferentes índices -sin dejar de dar crédito a los autores- denominaremos LL1 y LL2 a las fórmulas de López Morales-Lorán y LS a la fórmula de López Chávez-Strassburger Frías.

en torno a centros de interés, así como en cuanto al empleo de la lengua hablada.

En nuestro estudio diferenciamos diez centros de interés que se discuten en detalle en el Capítulo V. Empleamos una fórmula propuesta por M. Echevarría<sup>31</sup> para medir el grado de coincidencia en las respuestas de los sujetos para un mismo centro de interés.

En el Capítulo VI presentamos nuestras conclusiones y algunas sugerencias para investigaciones futuras.

6. Recuento bibliográfico. Nuestra investigación se refiere al tema de la disponibilidad léxica. Como ya expresáramos se trata de un concepto relativamente reciente; por lo tanto, existe un número limitado de trabajos relacionados directamente con él.

El primer estudioso en presentar el concepto fue Michéa<sup>32</sup> en 1950, cuando hace la distinción entre palabras "temáticas" y "atemáticas". Para Michéa las primeras se presentan en situaciones concretas, en tanto las segundas pueden presentarse con regularidad, no importa el tipo de texto que se analice. Las palabras "atemáticas" son para él las gramaticales, una serie de verbos y adjetivos, y algunos sustantivos muy usados. Estas serían las palabras frecuentes, en tanto las "temáticas" serían las palabras

---

31. M. S. Echevarría et al. Loc. Cit., p. 23.

32. R. Michéa, "Vocabulaire et culture", Les Langues Modernes, mayo-junio, 1950, pp. 188-189.

disponibles. En 1953 publica un artículo que titula "Mots fréquents et mots disponibles. Un aspect nouveau de la statistique du langage"<sup>33</sup>, en el que presenta su selección de centros de interés.

En 1959 se reedita la obra Le Français Élémentaire publicada en 1954 con el nombre de L'Elaboration du Français Fondamental (1er. degré)<sup>34</sup> con la innovación de tomar en consideración no sólo la frecuencia de los vocablos sino también la disponibilidad. Otra innovación es tener en cuenta el lenguaje hablado, puesto que hasta el momento se había trabajado únicamente con lengua escrita. El texto se propone como instrumento de difusión rápida de la lengua francesa y su interés es fundamentalmente pedagógico. En cuanto a la disponibilidad utilizan las variables geográficas, sociológicas y de sexo señalando que las sociológicas son relevantes, en menor grado la variante de sexo e irrelevante la zona geográfica.

Naum Dimitrijevic<sup>35</sup> publica en 1969 su Lexical Availability; se basa en los centros de interés pero, una variante importante de su trabajo es que mientras Michéa, Gougenheim, Rivenc y Sauvageot utilizan el criterio de listas cerradas de 20 palabras en sus encuestas, Dimitrijevic propone el criterio de listas abiertas con un

---

33. R. Michéa, Op. Cit.

34. G. Gougenhém et al., Op. Cit.

35. N. Dimitrijevic, Lexical Availability, Germany, Julius Groos Verlag, 1969.

tiempo fijo. Plantea su segmentación de unidades como lexical phrases, ya que las respuestas pueden constar de una, dos o más palabras. Aunque los estudios de vocabulario pueden dividir éste en productivo (el vocabulario que podemos utilizar tanto por escrito como oralmente) y receptivo (aquel vocabulario que podemos entender), su interés radica en encontrar la disponibilidad de palabras en once centros de interés del vocabulario productivo. Un factor interesante en las variables que considera en su estudio es la inteligencia de los encuestados. Su investigación se realiza en Edinburgo donde la población de escuela secundaria se divide en Senior Secondary School y Junior Secondary School dependiendo del coeficiente de inteligencia de los alumnos.<sup>36</sup>

En Canadá, W. Mackey et al.<sup>37</sup> intentan documentar una comparación del vocabulario común de dos regiones francófonas, se basan en la idea de disponibilidad y estudian dieciséis centros de interés. Su meta es hacer un estudio comparativo del léxico usado en Francia y de éste, qué palabras son conocidas o utilizadas en el francés canadiense. Para establecer el vocabulario común tienen en cuenta tres medidas: frecuencia, dispersión y valencia. Su

---

<sup>36</sup>. Dimitrijevic nos dice en Lexical Availability, p. 9. "... Intelligence of the pupils is one of the factors being taken into account when deciding which pupil will be taken to a senior secondary school or a junior secondary school. This means that the mere division of the pupils into those who attended SSS or JSS implies a division of their intelligence ability."

<sup>37</sup>. W. F. Mackey et al., Op. Cit.

concepto de valencia se refiere a las palabras con gran poder de combinación, que pueden incluirse en diversos textos y las denominan de extensión polivalente. Expresan que el estudio de las palabras concretas que reflejan las diferencias geográficas y culturales son las que realmente hacen la diferencia en el momento de la comparación. A estos conceptos agregan el de disponibilidad para los nombres concretos basándose en la idea de Michéa, y la toman como medida de diferencias culturales. Por ser su interés la comparación deben tomar los mismos centros de interés que Gougenhéim (aunque añaden cuatro más) y deben escoger la muestra entre estudiantes de la misma edad que la encuesta francesa. Calculan el índice de disponibilidad mediante el conteo del número de apariciones de una palabra dividiéndolo por el total de sujetos y multiplicándolo por cien.

Si siguiendo el modelo de Mackey encontramos el estudio realizado por M. J. Azurmendi<sup>38</sup>, sobre bilingüismo vasco/español, en el que se investigan quince centros de interés. Pide a los encuestados sustantivos y verbos en ambas lenguas. Propone un método de análisis multivariante como el más idóneo para estudios sociolingüísticos del bilingüismo.

Uno de los estudios pioneros de la disponibilidad léxica en español se realiza con escolares de la zona metropolitana de San Juan de Puerto Rico por H. López Morales. El

---

38. M. J. Azurmendi Ayerbe, Elaboración de un modelo para la descripción sociolingüística ...

estudio se lleva a cabo en 1973 y el propósito era hacer un examen sociolingüístico de disponibilidad. Se seleccionaron diez centros de interés, y la muestra se tomó en tres escuelas representativas de dos niveles socio-económicos. La disponibilidad atendía a tres grados: disponibilidad máxima, media y mínima. De este trabajo sólo se han publicado dos calas<sup>39</sup>.

También en Puerto Rico, como ya señaláramos anteriormente, se emplea el concepto de índice de disponibilidad, en un planteamiento que se caracteriza por su formalidad y rigurosidad matemáticas en el artículo de R. Lorán de 1983.<sup>40</sup>

A partir del concepto de índice de disponibilidad como una medida que considera frecuencia y orden a la vez, se realizan varias investigaciones que siguen ese criterio. Se utiliza una de las fórmulas de Lorán en la investigación que lleva a cabo B. Román<sup>41</sup> en Puerto Rico, con listas de palabras desiguales y analizando los mismos diez centros de interés investigados por H. López Morales.

En la misma línea de investigación, que utiliza la fórmula de López Morales y Lorán para listas cerradas,

---

39. H. López Morales, Dialectología y Sociolingüística: Temas Puertorriqueños, Madrid, Hispanova S.A., 1979, pp. 173-181.

40. R. Lorán, Un índice de disponibilidad léxica, Comunicación Técnica, Departamento de Matemáticas, Recinto Universitario de Mayagüez, Universidad de Puerto Rico, 1983.

41. B. Román, Disponibilidad léxica en escolares de Dorado, (tesina inédita), Río Piedras, Univesidad de Puerto Rico, 1985.

encontramos un estudio realizado en México por H. Justo Hernández<sup>42</sup>. Nos resulta de interés el hecho de que esta investigación se realiza para medir la disponibilidad en centros de interés referidos a colores, con un muestra de 56 hablantes (30 hombres y 26 mujeres). La investigación se llevó a cabo con listas cerradas de 10 palabras por centro y éstos fueron: ropa, zapatos, cosméticos, automóviles, anuncios comerciales, mochilas-maletas-bolsos, muebles, tapices-alfombras, pintura de casa (exterior-interior), vestido.

También en México A. V. Canizal Arévalo<sup>43</sup> realiza una investigación pero, utiliza la fórmula LL2 para listas abiertas o desiguales con un tiempo de respuesta limitado a cinco minutos por centro de interés. Limita el estudio a seis centros de interés que son: las partes del cuerpo, objetos que hay en el interior de la casa, alimentos y bebidas, medios de transporte, profesiones y oficios, defectos físicos y morales (adjetivos). Una innovación es que no sólo le interesan los nombres, sino que en un centro específico se tiene en cuenta particularmente los adjetivos. La muestra es de sesenta y ocho informantes y las variables que considera son sexo, escuela pública y privada, y cantidad de horas que los informantes ven televisión.

Un breve estudio sobre el español realizado en México por

---

42. H. G. Justo Hernández, Op. Cit.

43. A. V. Canizal Arévalo, Op. Cit.

R. Ramírez Gasca<sup>44</sup> analiza los anglicismos de uso cotidiano considerados los más frecuentes entre estudiantes de la Facultad de Filosofía y Letras de la Universidad Nacional Autónoma de México. Establece la disponibilidad para cada campo léxico semántico, considerando el número de conceptos respondidos por más del 50% de los informantes en relación con el número total de conceptos.

Estudiosos mejicanos se interesaron en el tema y fundamentalmente en la formalización para el análisis de los datos de disponibilidad léxica recogidos en México. J. López Chávez y C. Strassburger Frías.<sup>45</sup> plantean en un artículo de 1987 una nueva fórmula que denominan "Otro cálculo del índice de disponibilidad léxica".

Por otra parte, es interesante destacar que en Chile se realizan investigaciones paralelas a las que señaláramos en México. M. Mena Osorio<sup>46</sup> hace un estudio de disponibilidad basándose en una amplia muestra de 300 sujetos de cuarto, sexto y octavo año de enseñanza básica, y considera variables sociolingüísticas como estrato social y sexo, utiliza la fórmula LL2. Su análisis de los datos es cuantitativo y cualitativo, éste último centrado en el tipo de léxico emitido por los informantes.

---

44. R. Ramírez Gasca, "Uso de anglicismos en el habla estudiantil", Actas de II Congreso Internacional sobre el Español de América, México, UNAM, 1986, pp.589-592.

45. J. López Chávez y C. Strassburger Frías, Op. Cit.

46. M. Mena Osorio, Disponibilidad léxica infantil en tres niveles de enseñanza básica, (tesis inédita), Concepción, Chile, Universidad de Concepción, 1986.



Un estudio que nos resultó particularmente interesante fue el realizado recientemente por M. Echeverría et al.<sup>47</sup>, Disponibilidad léxica en Educación Media. Resulta novedosa su clasificación de centros de interés en: centros de alto grado de coherencia semántica y centros de bajo grado de cohesión semántica; propone un índice de cohesión, que da la medida en que los sujetos coinciden en la selección de sus respuestas en un centro en particular y además calcula el índice de disponibilidad mediante la fórmula LL1.

---

47. M. Echeverría, Op. Cit.

## CAPITULO II

### LOS INDICES DE DISPONIBILIDAD

1. Introducción. El problema de establecer un orden léxico que represente la facilidad o rapidez con que cada vocablo está accesible para un grupo de informantes se ha resuelto tradicionalmente mediante el uso de la frecuencia de menciones del vocablo en el grupo o una muestra de él. Como mencionáramos anteriormente, este enfoque es criticable principalmente por no tener en cuenta el orden o jerarquía que las palabras tienen para los informantes y que les conduce a mencionarlas o escribirlas unas antes que otras. En otras palabras, empleando la frecuencia solamente se está perdiendo información.

Los índices son medios para ordenar las unidades léxicas mediante fórmulas matemáticas. Estas, a partir de la frecuencia y el orden de mención de las palabras, entregan un número llamado índice de disponibilidad en base al cual se realiza la ordenación de las palabras correspondientes, de manera que palabras de mayor índice irán primero.

Los índices se han obtenido mediante fórmulas, deducidas rigurosamente, a partir de unas propiedades pre-establecidas para el ordenamiento, o bien, escogidas más o menos empíricamente mediante un proceso de proposición-prueba-rechazo... proposición-prueba-aceptación.

Cuando hablamos de ordenación en un conjunto de palabras, nos estamos refiriendo a que dadas dos palabras cualesquiera de ese conjunto, o bien una antecede a la otra o ambas están

en el mismo puesto. En terminología matemática nos estamos refiriendo a una relación de orden débil, es decir una relación que verifica las propiedades lineal<sup>1</sup> y transitiva<sup>2</sup>.

Nuestro corpus estará compuesto por listas de palabras provistas en cierto orden por informantes que responden a estímulos dados. Como es obvio una palabra cualquiera puede ser mencionada por más de un informante y, por tanto, es posible contar cuantas menciones ha tenido cada vocablo en cada posición.

Supongamos que un grupo de tres informantes produzca las siguientes listas:

	<u>sujeto 1</u>	<u>sujeto 2</u>	<u>sujeto 3</u>
	P1	P2	P1
	P3	P4	P2
	P2	P1	P4
	P5	P3	
		P5	

Al tamaño de cada lista le llamaremos rango del informante y éste no tiene por qué ser necesariamente constante. Nótese que en una posición fija no todas las palabras tienen igual número de menciones o frecuencias y que es posible construir para cada una de ellas un vector de

1. La propiedad lineal para una relación significa que dadas dos palabras cualesquiera o bien una de las dos antecede a la otra o ambas están en el mismo puesto.

2. La propiedad transitiva en una relación significa que si una palabra  $p_1$  antecede a otra  $p_2$  y ésta a su vez a otra  $p_3$  entonces  $p_1$  antecede a  $p_3$ .

frecuencias de la siguiente manera:

p <sub>1</sub>	2	0	1	0	0
p <sub>2</sub>	1	1	1	0	0
p <sub>3</sub>	0	1	0	1	0
p <sub>4</sub>	0	1	1	0	0
p <sub>5</sub>	0	0	0	1	0

Los números a la derecha de p<sub>1</sub> significan que dos veces fue mencionada esa palabra en primer lugar; que ningún informante la mencionó en segundo lugar; que sólo una persona la incluyó en tercera posición y que p<sub>1</sub> no tuvo posteriormente más menciones. Las frecuencias así dispuestas forman una matriz numérica a la que llamaremos matriz de frecuencias en que el número de filas es igual al de palabras diferentes provistas por todos los informantes y el número de columnas coincide con el mayor de los rangos y al que llamaremos rango de la encuesta.

Antes de proseguir es importante recalcar que lo que se busca obtener mediante los índices de disponibilidad es una alternativa al empleo de la frecuencia como medio de ordenar los vocablos provistos por un grupo de informantes. Como veremos, con cualquiera de los índices que consideramos podemos encontrarnos que si bien una palabra p puede haber sido mencionada más veces que otra p' el índice de disponibilidad puede ser mayor para la segunda que para la primera. Esto podría darse en el caso de que todas o gran parte de las menciones de p sean en posiciones bastante

posteriores que las menciones obtenidas para p'. En los resultados que presentamos en el Apéndice A incluimos para cada palabra su frecuencia a los efectos de poder comparar el orden que ésta genera con los órdenes representados por las diferentes fórmulas en discusión.

En la Sección 2 del capítulo anterior mencionábamos la fuerza asociativa y concretamente hacíamos referencia a la fórmula propuesta por Njock<sup>3</sup> para evaluar la fuerza con que un vocablo está asociado a una idea o centro de interés para un grupo de informantes. Es importante señalar la diferencia existente entre esta medida y los índices de disponibilidad que discutimos. Podría darse el caso que una palabra que recibió una sola mención sea anterior a otra que recibió muchas menciones por el solo hecho de que estas menciones son todas posteriores a la de la primera palabra. A pesar de que la asociación de una palabra a la idea propuesta o centro de interés es un concepto muy importante, éste es sólo un aspecto de la disponibilidad, al igual que lo es la frecuencia de menciones. Quizás podríamos decir que ambos son los principales elementos que los índices de disponibilidad tratan de tener en cuenta. En los resultados que presentamos en el Apéndice A, tenemos para cada palabra su fuerza asociativa calculada por la fórmula de Njock así como el orden correspondiente a los efectos de poder comparar la ordenación que este índice refleja con los otros órdenes presentados.

---

<sup>3</sup>. M. J. Azurmendi Ayerbe, Loc. Cit., p. 186.

En los cuadros que componen el Apéndice A tenemos listas de palabras que corresponden al centro de interés "alimentos", para las diferentes escuelas y para sexo masculino. En cada cuadro, que representa los valores de las variables escuela y sexo, a la derecha de cada palabra observamos:

a) en las columnas 1 y 2 respectivamente la frecuencia y el orden que correspondería a la palabra de ser éste el criterio empleado para la ordenación.

b) en las columnas 3 y 4 respectivamente el índice calculado mediante la fórmula LL1 calculado con valor .75 para la tasa de sustitución y el orden correspondiente.

c) en las columnas 5 a 10 los índices calculados con la fórmula LL2 con valores diferentes para la tasa de sustitución y los órdenes respectivos.

d) en las columnas 11 y 12 el índice calculado con la fórmula LS y el orden correspondiente.

e) en las columnas 13 y 14 respectivamente, la fuerza asociativa (fórmula de Njock) y el orden que correspondería a la palabra, según este criterio de ordenación.

A continuación discutiremos detalladamente los índices de disponibilidad obtenidos con LL1, LL2 y LS<sup>4</sup>.

2. El Índice LL1. El ambiente de aplicación para esta fórmula se caracteriza principalmente por unos listados todos del mismo número de palabras provenientes de un grupo

---

4. Ver nota número 30 en la pág. 24.

de informantes en respuesta a cierto centro de interés.

Aparte de haber sido casi con certeza la primera fórmula de este tipo, tiene el valor de su rigurosidad matemática. No es frecuente encontrar en lingüística, en los trabajos de orientación cuantitativa, un enfoque tan riguroso. Para obtener el índice se discute primeramente cuáles son las características que se espera de éste para de allí deducir cuál es la fórmula que debe emplearse. Estas características son llamadas axiomas y las discutiremos a continuación.

El primer axioma establece la llamada independencia de los componentes de los vectores de frecuencias relativas. Esto significa que si tenemos un par de vectores de frecuencias con algunos componentes comunes a ambos y si esos componentes comunes se cambiaran por otros, también comunes el orden de los vectores o -lo que es lo mismo- de las palabras que estos representan sigue siendo el mismo. Veamos un ejemplo. Consideremos las siguientes dos parejas de palabras con sus correspondientes vectores:

cons...	$p_1 : (0.10, 0.20, 0.33, 0.21)$
inform...	$p_2 : (0.11, 0.19, 0.33, 0.21)$
deter...	
	Y
como...	$p'_1 : (0.10, 0.20, 0.01, 0.42)$
de...	$p'_2 : (0.11, 0.19, 0.01, 0.42)$
exper...	

Nótese que la primer pareja tiene las posiciones tercera y cuarta iguales; lo mismo ocurre para la segunda pareja. De

acuerdo al primer axioma si  $p_2$  es anterior en la ordenación a  $p_1$  -lo cual parece ser razonable- deberá ser  $p'_1$  anterior a  $p'_2$ .

Uno de los conceptos de mayor importancia en la selección de la fórmula adecuada es el de tasa de sustitución. Este concepto si bien es aplicado por primera vez en lingüística en el trabajo de Lorán y López Morales, es de empleo frecuente en la Teoría de Decisiones.

Se llama tasa de sustitución de una posición respecto de la siguiente a la razón entre una variación en la frecuencia relativa en la primera posición y el cambio necesario en la siguiente posición para compensar la variación.

¿Qué valor debe darse a la tasa de sustitución?. Como se desprende de nuestras consideraciones sobre la frecuencia como medio de producir una ordenación léxica, es necesario dar mayor peso a las menciones en los lugares más adelantados que en los atrasados; por tanto la tasa de sustitución deberá ser necesariamente menor que la unidad.

El axioma segundo establece que la tasa de sustitución es constante. Aunque con este axioma ya tenemos más información sobre el valor de la tasa, ésta no está aún determinada; como veremos, la fórmula final la contendrá como un parámetro al que debe darse un valor.

Como método para su determinación se sugiere el empleo de expertos, quienes colocados frente a ordenaciones basadas en la fórmula LL1 y empleando diferentes valores para la tasa puedan decidir cuál es el valor más apropiado.



El tercer axioma recoge otra de las características que resulta deseable para las ordenaciones léxicas que se pretenden. Quizás a diferencia del anterior, su aceptación es menos controvertible y no requiere mucho esfuerzo. Este axioma dice que entre dos palabras con el mismo número de menciones excepto en dos posiciones, la más disponible será aquélla que haya sido mencionada más veces en el lugar más adelantado.

El último de los axiomas no tiene más propósito que establecer las condiciones para poder efectuar comparaciones entre diferentes ordenaciones o estudios. Este dice que una palabra mencionada siempre en primer lugar por todos los informante debe ser uno. Obviamente en este caso coincidiría el índice de disponibilidad con la frecuencia relativa clásica.

Basándose en diferentes resultados clásicos de la Teoría de Decisiones R. Lorán<sup>5</sup> demuestra que si los cuatro axiomas anteriores se verifican entonces la fórmula que representará esa ordenación y que llamamos LL1 es la siguiente:

con

$$d(p) = \sum_{i=1}^n \lambda^{i-1} x_{pi}, \text{ siendo } x_{pi} = \frac{f_{pi}}{N_i}$$

donde

Donde  $n$  es la máxima posición alcanzada; donde  $x_{pi}$  es la frecuencia relativa y  $f_{pi}$  es la frecuencia absoluta de la palabra  $p$  en la posición  $i$  y  $N_i$  es el número de informantes cuyas listas contenían la posición  $i$ .

---

<sup>5</sup>. R. Lorán. Loc. Cit., p. 22.

Para poder aplicar esta fórmula, que como dijimos requiere listados de igual tamaño, consideramos sólo parte de las listas provistas por los informantes. Concretamente, recortamos las listas para que todas tuvieran el largo de la más corta de ellas. En las tablas del Apéndice A, en las columnas 3 y 4 incluimos los valores del índice de disponibilidad (de las palabras de las listas recortadas) y el orden generado por este índice respectivamente.

Aunque posteriormente nos extenderemos en la comparación de las diferentes fórmulas es de señalar que no conocemos ninguna objeción a esta fórmula que es, además, de sencilla aplicación.

En cuanto a su aplicación para listados de diferente tamaño no vemos otra alternativa que la aquí empleada, de recortar las listas a la longitud del informante que respondió menos palabras. Obviamente la matriz de frecuencias considerada es una submatriz de aquélla que considera toda la información recogida.

Si bien en los casos de listados de diferente tamaño- como el de nuestra muestra- no se aprovecha toda la información disponible, su aplicación puede ser conveniente. Gougenhém<sup>6</sup> utilizó los listados cerrados en su encuesta con escolares limitándolos a 20 palabras, con un tiempo máximo para contestar de 15 minutos.

del 3. El Índice LL2. Con esta fórmula se pretende resolver

<sup>6</sup>. G. Gougenhém et al., Loc. Cit., p. 154.

el problema que aún dejaba pendiente la del índice LL1: ordenar los vocablos proveniente de listados de diferente tamaño.

Sus autores procuran que LL1 sea un caso particular de LL2 y para lograrlo, a la vez que quitan peso en el cálculo del índice a las menciones en posiciones adelantadas (últimas menciones), modifican levemente la tasa de sustitución. Continuando con el mismo enfoque axiomático empleado anteriormente reemplazan el segundo axioma anterior por uno nuevo que establece que la tasa de sustitución para la posición número  $i$  es de la forma:

$$\lambda_i = \lambda^{i-1} \frac{N_i}{N_{i-1}}$$

por lo que la fórmula LL2 adquiere la forma siguiente:

$$d(p) = \sum_{i=1}^n \lambda^{i-1} \frac{N_i}{N_{i-1}} x_{pi}$$

o bien

$$d(p) = \sum_{i=1}^n \lambda^{i-1} \frac{f_{pi}}{N_{i-1}}$$

Siendo  $N_0 = N_1$ . Nótese que siempre  $N_i \leq N_{i-1}$ , y por tanto la nueva tasa es menor o igual que la empleada en la fórmula anterior, obviamente si los listados fueran iguales sería  $N_i = N_{i-1}$  y por tanto la tasa de sustitución coincidiría con la del axioma 2 expuesto anteriormente.

Luego de analizar los resultados que mostramos en el

Apéndice A, tanto respecto del valor de LL2 para las diferentes palabras como para el ordenamiento correspondiente nos llamaron la atención, así como también lo señalan López Chávez y Strassburger Frías<sup>7</sup>, ciertas irregularidades que apuntaban hacia alguna deficiencia en el nuevo segundo axioma.

Luego de analizar las relaciones algebraicas que se derivan de un caso muy sencillo Lorán<sup>8</sup> encuentra la razón y la solución para dichas irregularidades. Supongamos dos palabras  $p_1$  y  $p_2$  con vectores compuestos por ceros excepto en la posición  $i$  y en la siguiente  $(i+1)$  como se indican:

$$p_1 = (0, \dots, 0, 1, 0, 0, \dots, 0)$$

$$p_2 = (0, \dots, 0, 0, 1, 0, \dots, 0)$$

obviamente debe ser:

$$d(p_1) > d(p_2) \quad \text{y por tanto} \quad \frac{d(p_2)}{d(p_1)} < 1$$

aplicando la fórmula LL2 tenemos:

$$d(p_1) = \frac{\lambda^{i-1}}{N_{i-1}} \quad \text{y} \quad d(p_2) = \frac{\lambda^i}{N_i}$$

7. J. López Chávez y C. Strassburger Frías, Loc. Cit., p. 6 en que dicen "... Todo depende del número de informantes y de la proporción que guarden en el descenso."

8. R. Lorán, La tasa de sustitución y el índice de disponibilidad léxica, Comunicación Técnica, Depto. de Matemáticas, Recinto Universitario de Mayagüez, Univ. de Puerto Rico, 1987.

de donde

$$\frac{\lambda N_{i-1}}{N_i} < 1 \quad \text{y} \quad \lambda < \frac{N_i}{N_{i-1}}$$

Como puede verse de la última desigualdad, si  $\lambda$  es 0.9 podría ocurrir que el término de la derecha no lo supere, esto es especialmente probable cuando la disminución de  $N_i$  respecto de  $N_{i-1}$  es brusca. Pero si se escogiera adecuadamente  $\lambda$  sería posible evitar este tipo de irregularidad, con lo cual la fórmula trabajaría perfectamente.

En el Apéndice A incluimos en las columnas 9 y 10 el cálculo de LL2 (para  $\lambda = 0.9$ ) y la posición resultante respectivamente para cada palabra, lo mismo en las columnas 7 y 8 para  $\lambda = 0.75$  y en las columnas 5 y 6 para  $\lambda = 0.60$ .

Nótese que sería posible determinar  $\lambda$  de manera que aún en las condiciones más desfavorables para  $N_i$  y  $N_{i-1}$  no apareciese ninguna irregularidad. El máximo valor que toma  $N_{i-1}$  en nuestro trabajo es de 19 y el mayor salto se produce cuando  $N_i$  es 11 y por tanto  $\lambda = 0.578$ .

Podría ocurrir que con esta forma de calcular  $\lambda$  no fuera posible emplear el método de vectores igualmente disponibles sugerido por Lorán en el Apéndice de su trabajo. En todo caso nos parece que ello no invalida la fórmula, ya que es posible encontrar un valor de compromiso entre las posibles contradicciones y la opinión de los expertos reflejada al establecer vectores que se desea sean

equivalentes en la ordenación resultante.

4. El Índice LS. En la sección anterior nos referíamos a eventuales irregularidades que sin una adecuada selección del parámetro  $\lambda$ , podrían ocurrir. López Chávez y Strassburger<sup>9</sup> proponen una nueva fórmula cuyo propósito principal es el evitar dichas situaciones.

La nueva fórmula consiste principalmente en el reemplazo de la tasa de sustitución empleada en LL1 y LL2. Su forma es la siguiente:

$$d(p) = \sum_{i=1}^n \text{EXP}\left(-c \frac{i-1}{n-1}\right) \frac{f_{pi}}{N_1}$$

Donde  $n$  es la máxima posición alcanzada,  $f_{pi}$  es la frecuencia absoluta de la palabra  $p$  en la posición  $i$ ;  $N_1$  es el número de informantes que participaron en la encuesta<sup>10</sup>.

Con respecto al valor de la constante  $c$ , el valor sugerido por los autores es  $c = 2.3$ . Este valor permite que los valores de la tasa de sustitución varíen desde 1 a 0.1 entre las primera y última menciones.

A diferencia de las fórmulas LL1 y LL2 ésta tiene una base empírica más que axiomática y los autores llegan a ella luego de realizar numerosas pruebas y evaluar los órdenes así obtenidos.

---

<sup>9</sup>. J. López Chávez y C. Strassburger Frías, Loc. Cit., p. 20.

<sup>10</sup>. También llamado  $I_1$  por los autores de la fórmula.

Son de mencionar los diferentes intentos realizados antes de llegar a LS1 y las razones porque éstos fueron rechazados. Un primer intento fue la fórmula siguiente:

$$d(p) = \sum_{i=1}^n \text{EXP}[-(i/n)^2] \frac{f_{pi}}{2N_1 - N_i}$$

Según señalan los autores ésta fue rechazada debido a que la tasa de sustitución variaba muy poco, de 0.99 a 0.40, dando valores para el índice muy poco dispersos. En adición, el rango de la prueba (n), intervenía de forma que los valores primero y último coincidían para cualquier n.

El segundo intento fue:

$$d(p) = \sum_{i=1}^n \text{EXP}[-2.3(i/n)^2] \frac{f_{pi}}{2N_1 - N_i}$$

Sobre esta propuesta preliminar López Chávez y Strassburguer informan que empleándola se realizaron varias tesis pero que se observaron -lo cual confirmamos con nuestros datos- diversas irregularidades, como por ejemplo, que dos posiciones en segundo lugar tengan mayor disponibilidad que una en primero y una en tercero, lo cual contradice el interés en que la tasa de sustitución sea decreciente.

Irregularidades del tipo de la señalada fueron resueltas con una nueva versión<sup>11</sup>:

$$d(p) = \sum_{i=1}^n \text{EXP}[-2.3(i/n)] \frac{f_{pi}}{2N_1 - N_i}$$

la cual fue reemplazada por la versión final, la mencionada al comienzo de esta sección.

Es interesante señalar que la fórmula del índice LS no presenta el inconveniente señalado para la del LL2<sup>11b</sup>. En las tablas del Apéndice A se presenta en la columna 11 el valor del índice LS y en la columna 12 el orden que a cada una correspondería si éste hubiese sido el orden escogido.

Es de señalar que los autores no hacen mención de la existencia o no de otras fórmulas que conduzcan a índices y ordenamientos equivalentes.

5. Comparación entre los diferentes índices. Nos parece importante realizar una comparación entre los diferentes índices, o equivalentemente entre las ordenaciones generadas por éstos. El propósito de la comparación es poder descubrir relaciones entre ellos, así como su proximidad con dos ordenaciones a las que podemos llamar "extremas": la

11. Para las dos palabras escogidas tenemos que

$$\frac{d(p_2)}{d(p_1)} = \text{EXP}(-2.3/n) * \frac{2N_1 - N_i}{2N_1 - N_{i+1}}$$

sabemos que  $\frac{2N_1 - N_i}{2N_1 - N_{i+1}} \leq 1$  y que  $0 < \text{EXP}(-2.3/n) < 1$ .

Por tanto  $\frac{d(p_2)}{d(p_1)} < 1$  independientemente de  $N_i$  y de  $n$ .

11b. Para las dos palabras escogidas tenemos que

$$\frac{d(p_2)}{d(p_1)} = \text{EXP}(-2.3/(n-1)) < 1 \text{ para todo } n \text{ distinto de } 1.$$



generada por la frecuencia absoluta que ignora completamente el orden en que las palabras fueron ofrecidas, y la inducida por el rango medio que ignora completamente el número de menciones de las palabras y sólo trabaja con su posición.

Una de las herramientas estadísticas empleadas más frecuentemente para medir la correlación entre rangos u ordenaciones es la debida al estadístico inglés Charles Spearman<sup>12</sup>. Dicha fórmula llamada "índice de correlación de Spearman" permite dadas dos ordenaciones de un mismo grupo de sujetos (en nuestro caso palabras) obtener una medida de su grado de relación. Un caso extremo, en donde el índice toma su valor máximo de 1, se presenta cuando las dos ordenaciones son coincidentes. El otro extremo, en que el índice alcanza su mínimo valor de -1, se presenta cuando se tienen dos ordenaciones totalmente invertidas. La situación intermedia, con valores próximos a 0, correspondería al caso en que ambas ordenaciones fueran completamente aleatorias sin ninguna coincidencia o inversión en los órdenes.

El índice de correlación de Spearman se calcula mediante la fórmula siguiente:

$$\rho = 1 - \frac{6(\sum d_i^2)}{N(N^2-1)}$$

---

12. Una explicación del método de Spearman puede encontrarse en la mayoría de los textos de estadística general. Por ejemplo: C. I. Chase, Elementary Statistical Procedures, New York, McGraw-Hill Book Company, 1984, p. 145.

Siendo  $d_i$  la diferencia entre las posiciones que la palabra  $p$  ocupa en las dos ordenaciones y  $N$  el número total de palabras que se están ordenando.

En el Apéndice B se presentan las matrices en las cuales cada elemento es el valor del índice de correlación de Spearman para los ordenamientos indicados. Cada tabla corresponde al centro de interés "alimentos", al sexo masculino y a una escuela según se indica en la parte superior de cada una.

Las principales observaciones que hemos realizado de las tablas del Apéndice B y que permiten relacionar las diferentes ordenaciones son las siguientes:

1. La casi inexistente correlación entre el orden correspondiente a la frecuencia absoluta y el inducido por los rangos medios.
2. La elevada correlación entre los órdenes inducidos por LL1 y LL2 en ambos casos para  $\lambda = 0.75$ . Para posibilitar la comparación se tuvieron en cuenta sólo las palabras incluidas en las listas recortadas. Esto es coherente con la afirmación de Lorán en el sentido de que un índice es un caso particular del otro.
3. La elevada correlación entre las ordenaciones generadas por los índice LS y LL2 (con  $\lambda = 0.75$ ). Esto, además, confirma nuestra afirmación de que una adecuada selección de  $\lambda$  puede resolver las irregularidades encontradas cuando  $\lambda = 0.9$ .

4. En general la ordenación generada por LL2 (con  $\lambda = 0.75$ ) está más correlacionada con la generada por LS que la ordenación inducida por LL2 ( con  $\lambda = 0.90$ ).

6. Procesamiento de los datos. Tradicionalmente era muy difícil manipular tanta cantidad de datos, especialmente en los casos de encuestas grandes, los computadores vinieron en auxilio de los investigadores pero, sin embargo, no todo estaba resuelto, ya que aún se hacía necesario conseguir el casi siempre escaso recurso de un experto en programación.

Además de la dificultad en conseguir tal clase de técnicos es de mencionar el aumento consiguiente en los costos de la investigación. Por otra parte no es siempre fácil de lograr el entendimiento necesario entre la persona que busca unos resultados sin conocer con cierta profundidad las posibilidades de la programación y aquella persona que, por el contrario, no tiene claro en qué medida van a ser necesarios cambios importantes en la programación una vez que se empiecen a obtener resultados numéricos.

La posibilidad de disponer los datos de las encuestas en forma de matriz nos sugirió la idea de emplear en nuestra investigación los programas conocidos como "hojas de cálculo electrónico" llamadas en inglés "spread-sheets". Estos programas de bajo costo y disponibles para cualquier computador personal son fácilmente manejables y su propósito es fundamentalmente el servir como herramienta para el manejo de matrices. Aunque nuestro trabajo fue enteramente

realizado con el programa LOTUS-123 otros programas similares como por ejemplo APPLEWORKS, VISICALC o MULTIPLAN podrían ser igualmente útiles.

A continuación explicaremos en detalle el procedimiento seguido y cual fue la manipulación de los datos que realizamos para obtener los resultados numéricos que presentamos.

En el Apéndice C incluimos un ejemplar de las hojas que empleamos para la recogida de los datos. En el Capítulo V nos extenderemos sobre las razones que nos llevaron a emplear los ocho centros de interés con que se corresponden las columnas en que se divide dicha hoja. Con respecto al número de líneas de cada columna, que es de 25, éste fue consecuencia de estudios pilotos conducidos tanto por nosotros mismos como por estudiantes del Instituto de Lingüística.

En el Capítulo IV nos extenderemos sobre las razones por las cuales se seleccionaron las escuelas y grupos escogidos y ya hemos explicado las razones por las cuales orientamos nuestro trabajo a los quintos grados.

El trabajo de campo consistió en visitar las escuelas y salones y luego de distribuir los formularios, dar a los estudiantes un tiempo limitado a 5 minutos para cada uno de los centros. El proceso estuvo precedido de una mínima explicación sobre lo que debían realizar. Las encuestas fueron todas realizadas entre los meses de abril y mayo de 1987.

Luego de una mínima preparación manual de los datos procedimos a registrarlos en archivos en una microcomputadora. Cada archivo contiene una matriz u "hoja de cálculo" en que la primera columna corresponde a las diferentes palabras mencionadas, la segunda columna corresponde al número de menciones que la palabra a la izquierda ha tenido en primer lugar, la tercera columna es igual que la segunda pero con las menciones en segundo lugar y así sucesivamente (ver ejemplo en Apéndice D). En otras palabras, cada uno de estos archivos contiene la matriz de frecuencias de las palabras mencionadas por un grupo de informantes en relación a un centro. El número de filas de la matriz es igual al número de palabras diferentes mencionadas y el número de columnas es de 26 ya que es 25 el número de líneas de los formularios.

Un aspecto importante desde el punto de vista práctico fue la nomenclatura escogida para los diferentes "hojas de cálculo". Los archivos creados (nos referimos a las matrices de frecuencia) totalizaron 120 y correspondieron a todas las combinaciones posibles de las 4 escuelas, los 8 centros de interés y el sexo (masculino, femenino e integrado). El nombre de cada archivo estaba compuesto por tres letras. La primera representaba la escuela:

A: Amalia Expósito  
L: La Merced  
M: Modesto Rivera  
S: Salvador Brau  
T: María Teresa Serrano

La segunda representaba el sexo:

F : Femenino  
 M : Masculino  
 T : Masculino y Femenino integrados

La tercera de las letras representa el centro de interés según la siguiente codificación:

A : Alimentos	T : Naturaleza
N : Animales	C : Colores
L : La casa	H : Cuerpo humano
R : Ropa	J : Juegos y entretenimientos

Así por ejemplo el archivo LMA.WK1 corresponde a los estudiantes de quinto año del Colegio La Merced, de sexo masculino y al centro de interés "Alimentos". Cuando usamos únicamente dos letras nos referimos a determinada escuela y centro de interés, omitimos la variable sexo.

Una vez disponibles los archivos incluimos en cada "hoja de cálculo" las fórmulas necesarias par calcular los índices: LL1 (con  $\lambda = 0.75$ ), LL2 con tres diferentes valores de  $\lambda$ , LS y por último la fórmula de Njock.

Esta etapa es la más delicada, requiere una mayor comprensión del programa así como de las fórmulas que van a emplearse. Si bien no ofrece mayores dificultades para una persona con un poco de práctica en el uso de "hojas de cálculo electrónico", se requiere una mínima comprensión del manejo algebraico. En cualquier caso, esta forma de trabajo es mucho más accesible que el desarrollo de programación ya que la ayuda especializada es mínima y la flexibilidad muy superior.

Con la facilidad de "sort" u ordenar del programa ordenamos sucesivamente las palabras según cada uno de los índices calculados y antes de proceder a cada reordenación indicábamos la posición correspondiente en una columna junto a cada índice.

Para disminuir el tamaño de las matrices, facilitando su manejo y evitando afectar el tiempo de respuesta del microcomputador, extrajimos de cada matriz las palabras, los valores de los índices y sus posiciones correspondientes. Esta extracción nos condujo a 120 nuevos archivos a los que dimos el mismo nombre que a la matriz original pero anteponiendo una T. Así el archivo T\_LTC.WK1 corresponde a los índices y las correspondientes ordenaciones de las palabras mencionadas por los estudiantes de quinto grado de la escuela La Merced, de ambos sexos y en relación con el centro de interés "colores". Dichos archivos fueron empleados para realizar diferentes cálculos como por ejemplo calcular la correlación de Spearman.

Es importante mencionar que el volumen de datos que manejamos nos requirió ciertas características especiales del equipo. Nuestra recomendación para un estudio similar es el emplear un procesador equivalente o superior al de un microcomputador tipo AT de la compañía IBM. Al menos 512 kilobytes de memoria y un disco duro son imprescindibles. Nosotros utilizamos un microcomputador NCR modelo PC 8 con 512 kilobytes de memoria y un disco duro de 30 megabytes.

Como es obvio en un estudio de mayores proporciones, el

equipo requerido puede ser bastante más poderoso que el que empleamos nosotros. Un minicomputador podría llegar a ser necesario pero en todo caso siempre podría recurrirse al empleo de "hojas de cálculo electrónicas".

El uso de hojas de cálculo electrónicas es una técnica que permite realizar cálculos complejos de manera sencilla y rápida. Estas herramientas son muy útiles para el análisis de datos y la toma de decisiones.

Por otro lado, el uso de minicomputadores puede ser necesario en algunos casos para manejar grandes volúmenes de datos y realizar cálculos más complejos que las hojas de cálculo.

En conclusión, el uso de hojas de cálculo electrónicas es una opción viable y económica para la mayoría de los casos, mientras que el uso de minicomputadores puede ser necesario en situaciones más complejas.

de: 100  
palestina  
constitución



### CAPITULO III

#### NUEVOS INDICES DE DISPONIBILIDAD

1. Introducción. El objetivo, que hasta el momento se ha perseguido con el empleo de los índices de disponibilidad, se ha circunscrito a ordenar un conjunto<sup>1</sup> de palabras reflejando la "disponibilidad" que cada una de ellas tiene para un grupo de informantes. En el capítulo anterior hemos discutido extensamente las fórmulas conocidas para calcular el índice de una palabra para un grupo de informantes.

En el presente capítulo extendemos el concepto anterior proponiendo ordenaciones entre centros de interés a base de las palabras provistas por un grupo de informantes. Para ello hacemos corresponder un número al conjunto de palabras de cada centro de interés provisto por el grupo. Dicho número se llama "índice del centro para el grupo" y es una representación del orden deseado.

Otra aplicación en la misma línea de pensamiento la obtenemos asignando a cada conjunto de palabras un número, que representa un orden entre los informantes, el cual refleje la disponibilidad que para el grupo tienen las palabras, que emplea cada informante en particular.

Por último, sugerimos una forma de ordenar los grupos de

---

1. Para evitar confusiones al referirnos al colectivo de informantes le denominamos "grupo" y al colectivo de palabras le llamamos "conjunto" (un centro de interés constituirá un caso particular de conjunto de palabras).

informantes de acuerdo a la disponibilidad de los informantes individuales.

Las posibles combinaciones entre el número de informantes y el número de palabras así como la notación sugerida para los índices en cada combinación se muestran en el siguiente cuadro:

	p a l a b r a	c e n t r o	t o c o n s t r u c t o r e s
informante	1 -	2 $d_i(c)$	3 -
	4 -	5 $d_c(i)$	6 $d(i)$
grupo de informantes	7 $D_g(p)$	8 $D_g(c)$	9 -
	10 -	11 $D_c(g)$	12 $D(g)$

Cuadro 1  
Diversos tipos de índices

La notación que empleamos consiste en una "d" minúscula para los índices correspondientes a una sola persona y "D" mayúscula cuando corresponde a un grupo de informantes. Entre paréntesis se incluye la variable correspondiente y como subíndice el parámetro que permanece fijo. La notación para las variables y parámetros es la siguiente: g, grupo de informantes; p, palabra; c, centro de interés; i, informante.

Varias de las fórmulas e índices que presentamos se basan en los discutidos anteriormente; nótese que los razonamientos y fórmulas son válidos indistintamente para los diferentes índices de disponibilidad presentados en el Capítulo II.

En términos del cuadro anterior podríamos resumir el contenido de este capítulo como un conjunto de definiciones de "disponibilidad" obtenidas mediante agregaciones y generalizaciones de otras realizadas previamente.

diversos

suma

era

de

frecuencia

que

negativa

el

capítulo

los

datos

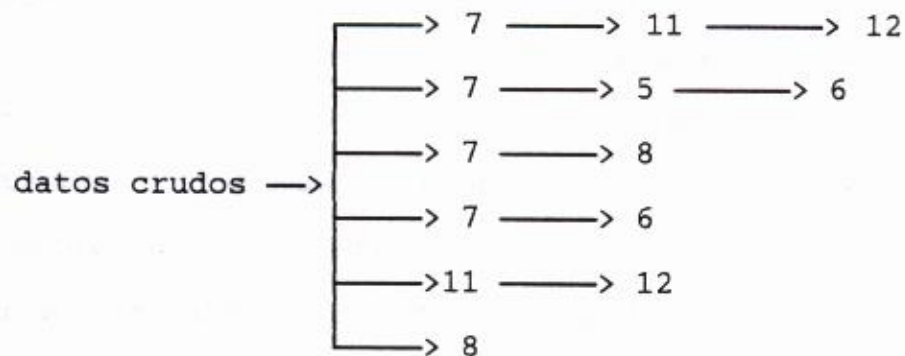
las

Nótese

para

procedimientos

los



Cuadro 2  
Posibles caminos para el cálculo  
de los diferentes índices.

El Cuadro 2 ilustra los caminos, que presentamos en este capítulo para calcular los índices de las diferentes casillas del Cuadro 1. Los caminos siempre se originan en los datos crudos, es decir tal y como fueron obtenidos de las encuestas. Las flechas indican sucesión de cálculos. Nótese que para llegar a las casillas 11, 12, 6 y 8 hay dos procedimientos posibles, mientras que solamente tenemos uno para llegar a las casillas 7 y 5.

de 2. Disponibilidad de centros para un grupo. Nos referimos ahora a la casilla 8 del Cuadro 1.

Para ordenar los centros de interés a base de las palabras suministradas en cada uno por un grupo fijo de informantes creemos que hay tres características que podrían ser tenidas en cuenta:

1) La variedad de palabras diferentes informadas por el grupo en el centro dado.

2) La frecuencia o número de menciones que las diversas palabras del centro han tenido para el grupo.

3) El orden en que estas palabras del centro fueron suministradas por los distintos informantes.

Cuando ordenábamos palabras, la primera característica no era relevante y las otras dos se reflejaban en los índices de disponibilidad. En el caso que nos ocupa, variedad y frecuencia son conceptos contradictorios en el sentido de que, una gran frecuencia de las palabras implica necesariamente un menor número de ellas y que, un aumento en el número de palabras traerá como consecuencia una disminución en la frecuencia de cada una. En nuestra encuesta constituye un ejemplo el conjunto LFC, en el cual el rango o total de palabras es de 33 con una frecuencia promedio de 6.4, mientras que en el conjunto LFA hay un total de 91 palabras con una frecuencia promedio de .30.

A continuación presentamos cuatro posibles índices de disponibilidad de un centro para un grupo ( $D_1$ ,  $D_2$ ,  $D_3$  y  $D_4$ ), los que pueden ser más o menos apropiados según sea el tipo

de estudio que se lleve a cabo.

Si sólo tenemos en cuenta la primera característica, el rango del grupo " $R_c$ " para ese centro nos permitiría establecer un orden entre los centros siendo:

$$D_{1g}(c) = R_c$$

Como es obvio, este índice depende directamente del número de informantes y tiene por tanto el inconveniente de no permitir comparaciones entre grupos diferentes.

El promedio del rango de los informante " $r$ " sería una forma de subsanar este inconveniente. A diferencia de la anterior esta medida depende de las habilidades de los individuos más que del grupo. Entonces sería:

$$D_{2g}(c) = \frac{\sum r_i}{I_c}$$

Siendo  $r_i$  el rango del informante  $i$ , e  $I_c$  el número de informantes que contribuyeron con palabras al centro  $c$ .

En el Cuadro 3 pueden observarse las medidas calculadas para el grupo femenino de la escuela "La Merced". En el caso del centro de interés "alimentos", el cuadro ilustra la diferencia entre ambas medidas; si bien la variedad de palabras informadas es la más alta, el rango promedio es bajo debido a aquellos informantes que presentaron listas breves.

Si sólo tenemos en cuenta el segundo criterio, la frecuencia promedio de las palabras lograría el mismo fin y

sería:

$$D_{3g}(c) = \frac{\sum f_p}{N_c}$$

Siendo  $f_p$  la frecuencia absoluta de la palabra  $p$  y  $N_c$  el número de palabras diferentes en el centro  $c$ .

Si dejásemos de lado la variedad de palabras la suma de disponibilidades de las palabras informadas sería un buen indicador. Nótese que una gran diversidad implica bajas disponibilidades y, por tanto, una suma pequeña:

$$D_{4g}(c) = \sum_p d_g(p)$$

Siendo  $d_g(p)$  el índice de disponibilidad de la palabra  $p$  del centro  $c$  para el grupo  $g$ , calculado mediante las fórmulas vistas en el capítulo anterior (LL1, LL2, LS, etc.).

En el Cuadro 3 también presentamos para el grupo femenino de la escuela "La Merced" los índices de los ocho centros de interés según las dos fórmulas anteriores y los correspondientes órdenes.

CENTRO	D <sub>1</sub>	D <sub>2</sub>	D <sub>3</sub>	D <sub>4</sub>
D(LF,A)	91 1	16.0 5	2.78 7	3.9600 5
D(LF,N)	83 2	19.2 1	3.39 5	3.9738 3
D(LF,L)	75 4	18.2 2	3.64 4	3.9800 2
D(LF,R)	49 7	14.6 6	4.53 2	3.9500 6
D(LF,T)	80 3	16.6 4	3.06 6	3.9736 4
D(LF,C)	32 8	14.5 7	6.59 1	3.9381 7
D(LF,H)	67 6	17.3 3	4.49 3	3.9903 1
D(LF,J)	68 5	11.2 8	2.52 8	3.8804 8

Cuadro 3  
Disponibilidad de grupo para diferentes  
centros según diferentes fórmulas

3. Disponibilidad léxica de un informante. Resultaría tanto de interés pedagógico y psicolingüístico, así como de valiosa herramienta en manos del lingüista, el poder disponer de una metodología para producir una ordenación del conjunto de los informantes basada en lo que podríamos llamar "disponibilidad léxica del informante". Este concepto sería un reflejo de la disponibilidad que dentro de un grupo dado tienen las palabras que emplea ( o de que "dispone") cada uno de los informantes.

Nos referimos ahora a las casillas 5 y 6 del Cuadro 1. Primero presentaremos la "disponibilidad léxica de un informante para un centro" (casilla 5) y posteriormente la "disponibilidad de un informante para todos los centros" (casilla 6).

Antes de proseguir es necesario aclarar que todas las afirmaciones o razonamientos que hacemos son realizados bajo el supuesto de que las condiciones y estímulos bajo los cuales se obtienen las respuestas de los informantes deben ser iguales. Cualquier diferencia invalidaría las comparaciones.

Aunque no pretendemos seguir un riguroso camino axiomático como el empleado para la fórmulas LL1 y LL2 creemos que existen algunas características que fácilmente podríamos establecer para un índice que refleje un ordenamiento como el que proponemos:

1. Supongamos que el informante I1 en cierta situación responde  $r$  palabras y que el informante I2, en las mismas circunstancias, responde  $s$  palabras, siendo  $r > s$  y que, además, todas las palabras mencionadas por I1 fueron mencionadas por I2. En esas circunstancias sería deseable que I1 tuviera una disponibilidad mayor que I2 y por tanto el índice del primero fuera mayor que el del segundo.

2. Supongamos que dos informantes, I1 e I2, responden igual número de palabras pero que las respondidas por I1 son en general más disponibles para el grupo que las respondidas por I2. En otras palabras, supongamos que las palabras ofrecidas por I2 son de poco o excepcional uso por parte del grupo. En este caso diríamos que I1 tiene una mayor disponibilidad léxica que I2.

Si la primera fuese la única condición requerida resultaría inmediato que para el caso de listas abiertas el



rango del informante, o sea, el número de palabras que cada informante ofrece, sería un posible índice. Pero creemos que deberían agregarse otras condiciones.

La segunda condición nos sugiere que el índice que deseamos para representar el orden entre informantes para un centro dado, debe depender también de la disponibilidad de las palabras ofrecidas. Si sumásemos la disponibilidad de todas las palabras ofrecidas por un informante para ese centro obtendríamos un índice que sería sensible a las dos situaciones mencionadas. Aumentaría la suma al aumentar el número de palabras y aumentaría o disminuiría la suma si aumentase o disminuyese respectivamente la disponibilidad de las palabras informadas. La fórmula de este índice sugerido tendría la siguiente forma:

$$d_c(i) = \sum_p D_g(p)$$

Siendo  $g$  el grupo al que pertenece la persona  $i$ ,  $p$  una palabra del centro  $c$  mencionada por la misma persona  $i$  y  $D_g(p)$ , el índice de disponibilidad de la palabra  $p$  perteneciente al centro  $c$  para el grupo  $g$ , calculado mediante las fórmulas vistas en el capítulo anterior (LL1, LL2, LS, etc.).

Como ya expresamos, los razonamientos y fórmulas que discutiremos a continuación son independientes de la fórmula que se escoja para calcular el índice de disponibilidad de las palabras.

En el Apéndice E incluimos como muestra algunos

formularios que contienen las palabras ofrecidas por informantes de sexo masculino de la escuela La Merced en los diferentes centros de interés. El Apéndice F ofrece para cada informante del mismo grupo y para cada centro de interés, en la segunda columna su rango y, en la siguiente, el orden generado por éste. En la cuarta y sexta se incluye el índice de disponibilidad de cada informante calculado con la fórmula anterior y a base de las fórmulas LL2-0.75 y LS y, en la quinta y séptima los órdenes respectivos inducidos por dichos índices.

Generalizaremos ahora los conceptos y fórmulas anteriores para el caso en que consideremos todas las palabras provistas por un informante, es decir todos los centros de interés. De forma inmediata se nos ocurre que una suma a través de todos los centros de la disponibilidad del individuo para un centro nos proveería de un índice adecuado. La fórmula sería entonces la siguiente:

$$d_1(i) = \sum_c d_c(i)$$

Siendo  $d_c(i)$  el índice del informante  $i$  para el centro  $c$ .

No obstante que las pruebas realizadas nos conducen a resultados razonables, encontramos que en algunas situaciones no se emplea toda la información que tenemos. Es el caso del informante con habilidad para incluir más palabras en centros en que el grupo completo no incluyó muchas palabras y que no tiene ningún impacto en el cálculo. Nos preguntamos si a los efectos de ordenar a dos

informantes debemos considerar todos los centros con igual importancia o peso.

Esta consideración nos sugiere una nueva fórmula en la cual, si bien se totalizan las disponibilidades de las palabras dadas por cada informante, se hace por centros separados y luego, antes de efectuar la suma para todos los centros, se da más peso a aquellos centros con pocas palabras. Aunque probamos diferentes formas de reforzar el peso de los centros más "difíciles" y amortiguar el de los "fáciles", encontramos que la más adecuada era emplear la razón entre el promedio de palabras mencionadas y el número de palabras de cada centro. Multiplicando la disponibilidad total de un centro con más palabras que el promedio estaremos restándole importancia al mismo y si el centro tiene menos palabras que el promedio estaremos realzando el centro. Nótese que estas razones no son directamente dependientes del informante sino que dependen solamente de la respuesta total del grupo en cada centro de interés. La fórmula que proponemos tiene la ventaja de su fácil aplicación y puede escribirse de la forma siguiente:

$$d_2(i) = \sum_c \frac{\bar{N}}{N_c} d_c(i)$$

Siendo  $d_c(i)$  el índice de disponibilidad del individuo  $i$  para el centro  $c$ ,  $N_c$  el número de palabras mencionadas en ese centro por el grupo completo, y  $\bar{N}$  el tamaño promedio de los centros.

En el Apéndice G para cada informante, en las columnas dos y cuatro incluimos el cálculo de las dos fórmulas anteriores empleando para las palabras el índice LL2-0.75; en las tercera y quinta incluimos la ordenación que correspondería a cada informante si aquéllos hubiesen sido los criterios de ordenación.

En forma muy escueta López Chávez y Strassburguer<sup>2</sup> informan que están trabajando en la dirección que aquí proponemos; hasta ahora, sin embargo, se han limitado a transcribir una fórmula, sobre la que no hemos podido obtener más información. La fórmula es la siguiente<sup>3</sup>:

$$d_3(i) = 1000 \sum_c \sum_{j=1}^{r_c} \text{EXP}[-2.3 \left( \frac{j-1}{n_c-1} \right) D(g,p)]$$

Siendo  $g$  el grupo al cual pertenece el individuo  $i$ ,  $p$  una palabra mencionada por el individuo en el centro  $c$ ,  $r_c$  el rango del individuo para el centro  $c$  y  $j$  es un indicador de la posición de la palabra en el centro correspondiente.  $D(g,p)$  representa aquí el índice de disponibilidad de la palabra  $p$  calculado mediante la fórmula LS.

4. Disponibilidad léxica de un grupo. El poder calcular la disponibilidad léxica de un grupo con el fin de realizar

---

2. J. López Chávez y C. Strassburger Frías, Op. Cit., p. 23.

3. Cambiamos levemente la notación empleada por los autores para poder mantener la consistencia con la empleada aquí, y facilitar así la tarea del lector.

comparaciones con variables sociolingüísticas nos parece que también podría representar un paso significativo.

Al igual que en situaciones anteriores es necesario que las condiciones o estímulos que afecten a todos los informantes del grupo sean idénticos.

Nos referimos ahora a la casilla 11 del Cuadro 1. Para ordenar los individuos de un grupo en la Sección 3, caracterizábamos a cada uno de éstos por un conjunto de palabras y luego, mediante una agregación de la disponibilidad léxica de dichas palabras, llegamos a la de los individuos. A primera vista, la situación que se nos plantea al tratar de ordenar los grupos de informantes parece similar. Al estar los grupos compuestos por personas y partiendo de la disponibilidad léxica de éstas, se podría llegar a la de los grupos mediante alguna forma de agregación de sus índices. Sin embargo, si examinamos con detenimiento esta situación, el planteamiento resulta inadecuado, dado que las disponibilidades de grupos diferentes nos son comparables por haber sido calculadas a base de diferentes datos.

Una solución que nos parece plausible sería el reunir toda la información suministrada por los grupos que queremos comparar, formando un conjunto de palabras más amplio sobre el cual hacer los cálculos. Luego de obtenida la disponibilidad de todas las palabras informadas por los grupos, se estaría en condiciones de calcular la disponibilidad para cada grupo por agregación de las

disponibilidades<sup>4</sup> de las palabras informadas por cada grupo en específico. Coherentemente la fórmula para calcular la disponibilidad de un grupo para un centro dado sería:

$$D_{1c}(g) = \sum_{p \in c} D_G(p)$$

Siendo  $D_G(p)$  la disponibilidad de la palabra  $p$  para el grupo total  $G$  y  $p$  una palabra mencionada por algún miembro del grupo  $g$  en el centro  $c$ .

Otro enfoque -que empleamos exitosamente- podría ser el considerar el conjunto de palabras provistas por cada grupo y a base de ellas calcular índices de disponibilidad para el grupo completo. Así, por ejemplo, podríamos utilizar:

$$D_{2c}(g) = \frac{R_{gc}}{I_{gc}}$$

Siendo  $R_{gc}$  el rango del grupo de informantes o número de palabras diferentes correspondientes al centro, e  $I_{gc}$  el número de informantes.

Otra alternativa<sup>5</sup> sería:

$$D_{3c}(g) = \frac{\sum r_{ic}}{I_g}$$

Siendo  $r_{ic}$  el rango del informante o número de palabras

---

4. Nos referimos al cálculo del índice de disponibilidad mediante el empleo de alguna de las fórmulas presentadas en el Capítulo II.

5. Esta alternativa la empleamos para calcular la disponibilidad de los grupos en el Capítulo IV.

diferentes informadas por éste en el centro  $c$ , e  $I_g$  el número de informantes que participaron en la encuesta. Una tercera alternativa sería:

$$D_{4C}(g) = \frac{\sum f_{pc}}{N_C}$$

Siendo  $f_{pc}$  la frecuencia absoluta de cada palabra del centro  $c$ , y  $N_C$  el número de palabras diferentes en dicho centro.

Una generalización del concepto de disponibilidad de un grupo para un centro dado, puede hacerse mediante agregación de las disponibilidades de todos los centros así:

$$D_5(g) = \sum_C D_C(g)$$

Una objeción a este tipo de fórmula podría ser el de la importancia relativa de los centros. Esta podría salvarse agregando algún tipo de ponderación, como en la fórmula siguiente:

$$D_6(g) = \sum_C w_C D_C(g)$$

Siendo  $w_C$  el peso correspondiente al centro  $c$ , y sea:

$$\sum w_C = 1$$

Una posibilidad para el cálculo de  $w_C$  fue dada en la Sección 3 haciendo<sup>6</sup>:

$$w_C = k \frac{N}{N_C}$$

Donde  $k$  es constante para que los pesos sumen 1.

Como puede deducirse, nuestro propósito en este capítulo es plantear el mayor número de posibilidades razonables, para el cálculo del índice de disponibilidad de un informante o de un grupo de informantes. Consideramos que cada investigador, según las características de su trabajo, puede encontrar una u otra forma de mayor utilidad.

---

<sup>6</sup>. Esta ponderación será empleada en un ejemplo en la Sección 5 del Capítulo IV.



CAPITULO IV  
ASPECTOS SOCIOLINGUISTICOS

1. Introducción. Desde prácticamente sus inicios, la llamada escuela estructuralista norteamericana -con L. Bloomfield a la cabeza<sup>1</sup>- había rechazado todo tipo de consideración sociológica en la confección de sus modelos y, consecuentemente, en el desarrollo de sus metodologías analíticas.

Los lingüistas europeos, por el contrario, creían -desde F. de Saussure<sup>2</sup>- en la naturaleza social del lenguaje, de modo que fue posible ver nacer en Francia el trabajo de un grupo de lingüistas que ahora la historiografía agrupa bajo el rubro de "escuela sociológica francesa" (Meillet, Vendryes, etc.).

Es cierto, sin embargo, que los factores sociales nunca pasaron a integrar la base misma de ningún modelo teórico hasta que en 1969 W. Labov sentaba las bases para ello<sup>3</sup>. Nació entonces la moderna sociolingüística.

Ya nadie pone en duda que cada sociolecto difiere en alguna medida de los otros. Por tal motivo en el plano léxico, aún en mayor grado, sería imposible hablar de

---

1. L. Bloomfield, Language, New York, Holt, Rinehart and Winston, 1961, (1era. ed. 1933).

2. F. de Saussure, Curso de lingüística general, Buenos Aires, Losada, 1945.

3. W. Labov, Sociolinguistic Patterns, Filadelfia, University of Pennsylvania Press, 1972.

disponibilidad léxica sin hacer referencia a unas variables sociolingüísticas. Por otra parte, los estudios de disponibilidad léxica aparecen como valiosa herramienta para los estudios sociolingüísticos.

Todos los estudios de disponibilidad léxica que hemos consultado parten de unas estratificaciones comunes: geográficas, de sexo, nivel socioeconómico, edad, etc..

El factor socioeconómico parece ser el factor extralingüístico más importante en la motivación de la variación lingüística. Muchos investigadores lo han tenido en cuenta en sus investigaciones al trabajar con disponibilidad léxica así, por ejemplo, Dimitrijevic<sup>4</sup>, López Morales<sup>5</sup> y Azurmendi<sup>6</sup>.

La variable geográfica se ha considerado de relativamente poca importancia, pero ha sido tomada en cuenta por autores como Mackey<sup>7</sup> pues su estudio se basa en una comparación regional. Gougenhém<sup>8</sup> encuentra diferencias entre lo que cataloga como medio urbano y medio rural.

Todos los estudios de disponibilidad coinciden en señalar que las diferencias que se aprecian en cuanto a la variable sexo son mucho más significativas que las diferencias con otras variables, especialmente en algunos centros de

---

4. N. Dimitrijevic, Loc. Cit., p. 11.

5. H. López Morales, Loc. Cit., p. 173.

6. M. J. Azurmendi, Loc. Cit., p. 351.

7. W. F. Mackey, Loc. Cit., p. 163.

8. G. Gougenhém, Loc. Cit., p. 165.

interés.

En nuestra encuesta tomamos en consideración la variable socioeconómica, la variable sexo y la geográfica. Empleando algunos de los índices presentados en los dos capítulos anteriores trataremos de verificar la relación de nuestras conclusiones con la realidad ya estudiada, haciendo énfasis en la validez de los índices como herramienta de investigación sociolingüística.

2. Lengua hablada; lengua escrita. Los estudios léxicos, dependiendo del interés del investigador, pueden tomar en cuenta la lengua hablada o la lengua escrita y pueden llevarse a cabo mediante cuestionarios, reconocimientos, análisis de significado, etc.. La mayoría de los estudios que conocemos hasta el momento han tomado en consideración la lengua escrita.

En la investigación llevada a cabo por Gougenhém se constatan las diferencias fundamentales que, en cuanto a léxico se refiere, se reflejan entre lengua hablada y lengua escrita, en especial en aquellos términos que pueden considerarse sinónimos o quasi sinónimos<sup>9</sup>.

---

9. Ibid., p. 118 nos expresa: "...Entre plusieurs mots synonymes ou quasi-synonymes, celui qui est le plus employé dans la langue parlée ressort avec vigueur, tandis que dans la langue littéraire les fréquences tendent à se rapprocher, d'une part en raison d'un maniement plus raffiné de la langue, d'autre part parce que les écrivains cherchent à introduire dans leurs écrits une variété dont la langue parlée se soucie peu."

Como muy acertadamente plantea Dimitrijevic<sup>10</sup> los resultados obtenidos en las investigaciones sobre léxico que han utilizado lengua escrita son muy válidos para ésta, pero no pueden aplicarse para un análisis objetivo de la lengua hablada. Los estudios realizados sobre lengua hablada son mucho menos numerosos. El autor señala en su libro que su investigación pretende ser un indicador de la fluidez oral de los informantes.

Es sabido que la rapidez de la producción oral es mucho mayor que la rapidez de la producción escrita, sin embargo, tratándose en nuestra muestra de estudiantes de quinto grado consideramos que la diferencia entre fluidez oral y escrita podría disminuir, si se les otorga a los niños un tiempo prudencial como para que la escritura no resulte un impedimento de importancia.

3. La encuesta. El interés primordial es aplicar a un corpus determinado algunas de las fórmulas presentadas en el Capítulo II, así como alguna de las sugeridas en el Capítulo III. Para tomar la muestra decidimos considerar las variables: socioeconómica, de sexo y geográfica.

Al no estar centrado nuestro interés en un estudio socioeconómico para llegar a una estratificación social lo más adecuada posible y a sabiendas de que nuestra elección en este aspecto no nos permitiría llegar a conclusiones definitivas, escogimos varias escuelas basándonos en la

---

10. N. Dimitrijevic, Loc. Cit., p. 3.

clasificación del Departamento de Instrucción Pública.

Los elementos en que se basa dicha clasificación en estratos socioeconómicos son: a) las barriadas donde se sitúan las escuelas y b) los tipos de familia de donde proviene el alumnado. La clasificación consta de tres niveles: alto, medio y bajo.

Las escuelas escogidas en la zona metropolitana de San Juan fueron: Amalia Expósito, clasificada como de nivel socioeconómico bajo; Dr. Modesto Rivera, de nivel socioeconómico medio; María Teresa Serrano, de nivel socioeconómico alto.

La variable sexo fue considerada en todos los grupos, ya que los estudiantes debían colocar en cada hoja de contestaciones una F o una M correspondiendo a su sexo.

Consideramos que la variable geográfica en algunos centros de interés podía tener alguna relevancia, siguiendo el criterio de Gougenhém; por eso, además de la zona metropolitana tomamos una muestra en la zona de Cayey. Consideramos ésta como representativa del centro de la Isla; además, es la zona donde residimos y habíamos notado alguna diferencias léxicas con otras partes de la Isla. Escogimos en Cayey una escuela pública clasificada como de nivel socioeconómico medio por el Departamento de Instrucción Pública, la escuela Salvador Brau. Escogimos una escuela privada para representar el nivel socioeconómico alto, la escuela La Merced.

En cada escuela se seleccionó un quinto grado. En las

escuelas en que había varios quintos grados, separando los estudiantes por niveles de rendimiento, escogimos el nivel intermedio, ni los considerados más aventajados, ni los que presentaban ciertos retrasos en lectura y/o escritura.

La encuesta se aplicó a toda la población, que en total ascendió a 154 estudiantes repartidos de la siguiente forma:

Amalia Expósito : 14 niñas y 15 niños  
Dr. Modesto Rivera: 20 niñas y 16 niños  
María Teresa Serrano: 20 niñas y 9 niños  
La Merced: 14 niñas y 13 niños  
Salvador Brau: 17 niñas y 15 niños

A los estudiantes se les entregaron unas hojas o formularios de contestación como el que incluimos en el Apéndice C. En reuniones preliminares con cada uno de los maestros de español explicamos los objetivos de la encuesta, el material que buscábamos, así como la forma de aplicación. El maestro de cada grupo fue el encargado de llevar a cabo la encuesta; aunque estuvimos presentes durante su realización, pretendíamos mantener el ambiente lo más natural posible en busca de un mayor grado de espontaneidad.

Los alumnos recibieron la explicación de como realizar el trabajo de su propio profesor que les indicaba:

1.- Deben identificarse únicamente por su sexo mediante la letra F para las niñas y la letra M para los niños. Deben colocar una "X" en el espacio correspondiente.

2.- Comenzarán a escribir a partir del número 1 de la columna titulada "Alimentos" desde el momento en que se dé la orden de comenzar y dejar de escribir en el momento que se indique. Así

sucesivamente para cada columna, donde deben colocar todas las palabras que se les ocurran en relación con el título.

El tiempo se limitó a 5 minutos para cada centro de interés, lo que sumando los ocho centros nos da un total de 40 minutos para llenar la hoja. Todas las encuestas se realizaron en el horario que correspondía a la clase de español.

4. Listas abiertas y listas cerradas. En su estudio de disponibilidad Gougenhém se propuso conocer el vocabulario más usado por los informantes y por tal motivo la encuesta se realizó con listas cerradas. Es decir que cada informante debía responder un número limitado de palabras (20) en un tiempo máximo de quince minutos.

Dimitrijevic, por su parte, se propone en su investigación descubrir qué palabras son las más disponibles y también cuántas palabras tienen disponibles los informantes. Es por tal motivo que no limita el número de palabras sino únicamente el tiempo de reacción. Considera que de tal forma y, facilitándoles un tiempo razonable para sus respuestas, se elimina el riesgo de contar con listados de palabras que rara vez acudirían a la mente del hablante en una situación normal y que sólo estarían disponibles luego de una meditación larga.

En nuestra investigación y con el propósito de comprobar el funcionamiento de los diferentes índices, consideramos que lo más adecuado era tener listas abiertas pues, como ya

señaláramos, podíamos probar la fórmula LL1 cortando las listas según el informante que respondiera menos palabras.

Como pensamos que la comparación de los índices de disponibilidad en listados iguales y desiguales podría arrojar alguna nueva idea sobre el problema de la encuesta óptima, en el Apéndice B calculamos el índice de correlación de Spearman entre las ordenaciones generadas por LL2-0.75 en listas desiguales y la misma fórmula<sup>11</sup> en el caso de las listas recortadas. Comprobamos que las ordenaciones están, en general, altamente correlacionadas y las diferencias se deben a las palabras que desaparecen al recortar las listas.

5. Comparaciones entre grupos. El interés de nuestra investigación no es, como se ha podido apreciar, de índole fundamentalmente sociolingüística. Nuestra muestra no es necesariamente representativa y por tanto las conclusiones son de alcance limitado. Sin embargo, con el propósito de visualizar el tipo de inferencias que podrían hacerse aplicamos la fórmula  $D_3C(g)$  presentada en la Sección 4 del Capítulo III.

En el Cuadro 4 pueden observarse los resultados para las cinco escuelas en los ocho centros de interés.

---

11. Cuando las listas son iguales LL2 coincide con LL1 para cualquier valor de  $\lambda$ .



	A	N	L	R	T	C	H	J
S	12.96	19.72	17.62	9.19	7.50	12.53	12.53	9.64
L	16.31	18.78	18.89	16.78	14.85	20.28	11.99	14.89
M	22.74	24.66	24.41	20.91	22.30	21.94	24.72	23.09
A	22.23	23.89	22.37	9.51	9.65	11.93	13.75	11.73
T	22.92	23.79	23.98	18.47	17.44	18.52	22.34	21.41

Cuadro 4  
Cómputo de la fórmula  $D3_C(g)$  para  
las escuelas en cada centro

En general se observan grandes diferencias en la misma escuela en los diferentes centros. La escuela más estable respecto a los centros de interés es la Modesto Rivera. Extrañan los resultados de la escuela Amalia Expósito en los centros "ropa" y "naturaleza".

En los cuadros 5 y 6 aplicamos la misma fórmula para los niños y las niñas por separado en las mismas cinco escuelas y para los diferentes centros de interés.

	A	N	L	R	T	C	H	J
SM	12.42	19.46	17.40	9.33	6.80	11.40	12.06	10.26
LM	15.69	17.92	19.23	16.23	14.62	19.92	12.46	14.85
MM	23.81	24.69	25.00	21.50	22.69	22.62	25.00	23.40
AM	21.73	22.87	20.80	9.87	10.27	10.47	13.27	10.33
TM	21.77	23.00	23.50	15.30	15.00	16.80	19.10	17.10

Cuadro 5  
Cómputo de la fórmula  $D3_C(g)$  para  
las escuelas (grupos masculinos)  
en cada centro de interés

	A	N	L	R	T	C	H	J
SF	13.12	19.94	17.82	9.06	8.12	13.53	12.94	9.11
LF	16.86	19.53	18.60	17.27	15.06	20.60	11.60	14.93
MF	21.90	24.65	23.95	20.45	22.00	21.40	24.50	22.85
AF	22.78	25.00	24.07	9.14	9.00	13.50	14.28	12.36
TF	23.45	24.15	24.20	19.90	18.55	19.30	23.80	23.35

Cuadro 6  
Cómputo de la fórmula  $D3_C(g)$  para  
las escuelas (grupos femeninos)  
en cada centro de interés

Coincidiendo con los numerosos estudios disponibles, la variable sexo resultó relevante y las diferencias entre los índices calculados para niños y niñas, en nuestra muestra, resultaron más perceptibles en algunos centros que en otros.

Lo que más llamó nuestra atención y que nos sugiere futuras aplicaciones de las fórmulas es la variable geográfica, que como ya señaláramos en la Sección 3, habíamos percibido subjetivamente. Puede observarse que la zona de Cayey, independientemente del nivel social o el sexo, está en gran desventaja con la zona metropolitana.

Un ejemplo de como pueden combinarse los índices calculados para los diferentes centros obteniendo uno común para todo el grupo, lo conseguimos tomando como insumo los datos del Cuadro 4.

Empleando una ponderación del tipo sugerido en la Sección 4 del Capítulo III tenemos:

$$w_A = 0.18 \quad w_N = 0.15 \quad w_L = 0.14 \quad w_R = 0.09$$

$$w_T = 0.14 \quad w_C = 0.06 \quad w_H = 0.12 \quad w_J = 0.14$$

obtenemos así los siguientes resultados:

escuela	D6(g)
S	12.99
L	16.39
M	23.25
A	16.64
T	21.49

Cuadro 7  
Cómputo de la fórmula D6(g) para  
las diferentes escuelas

Como puede observarse el nivel socio-económico es de gran relevancia. En nuestro caso, sin embargo, notamos que la diferencia fundamental se presenta entre los niveles extremos, resultando muy cercanos los niveles alto y medio. El nivel medio resulta superior al alto, lo que nos indica que debemos verificar los datos con una población más representativa y revisar los parámetros clasificatorios del Departamento de Instrucción Pública.

No creemos que quepa ninguna duda de que las comparaciones sociolingüísticas que puedan llevarse a cabo mediante el análisis de los resultados numéricos arrojarán luz sobre los múltiples problemas y carencias que presentan nuestros estudiantes, en cuanto a vocabulario se refiere.

## CAPITULO V

### CENTROS DE INTERES

1. Introducción. Este capítulo está dedicado principalmente, a exponer los elementos básicos relacionados con los centros de interés, así como los elementos que nos condujeron a la selección de centros para esta investigación.

En la Sección 4 desarrollamos los criterios de segmentación que nos condujeron al concepto de "palabra" que empleamos en la misma.

Por último en la Sección 5 discutimos el concepto de índice de cohesión propuesto por M. Echeverría y lo aplicamos a nuestros datos, esto nos permite algunas conclusiones sobre los centros de interés que empleamos.

2. Antecedentes. Luego de que Michéa planteara la importancia de estudiar aquellos nombres temáticos o concretos que sólo acceden a la mente del hablante en situaciones específicas, la pregunta obvia fue a qué situaciones específicas hay que enfrentar al hablante para conocer su vocabulario disponible.

La noción de disponibilidad corresponde a la presencia más o menos inmediata en la memoria de las palabras correspondientes a un campo semántico.

La expresión centro de interés designa, para la pedagogía tradicional, grupos que ofrecen una selección más o menos

numerosa de palabras. Se trata a veces de clases de objetos como por ejemplo, alimentos, animales, etc.; otras veces de ambientes: la casa, la escuela; también puede tratarse de situaciones cotidianas: una fiesta, una excursión etc.. Permiten al hablante derivar hacia un número de direcciones mayor o menor, es decir, puede haber una coherencia semántica diferente. Con el término coherencia semántica nos referimos a la limitación de clases lógicas que componen el centro. Es normal que las probabilidades de aparición de las clases no se jerarquicen exactamente de la misma forma en los diferentes centros. Con éstos se pretende realizar una exploración del inconsciente lingüístico, y su determinación será mejor toda vez que se logre un grado mayor de coherencia semántica.

Los centros de interés buscan el léxico corriente, el no especializado, el más utilizado en la vida diaria, el más ligado a las áreas significativas que se suponen más universales.

La disponibilidad es según Mackey<sup>1</sup> una medida de las diferencias culturales. El autor señala que la importancia del vino en Francia no es la misma para aquellas regiones vitivinícolas, que para las que no lo son.

En la primera encuesta que Gougenheim<sup>2</sup> realizó en Francia en 1954, basada en centros de interés, éstos representaban los sectores semánticos más universales a juicio del

---

1. W. F. Mackey, Loc. Cit., p. 26.

2. G. Gougenhém, Loc. Cit., p. 152.

estudioso que eran:

- 1) Les parties du corps
- 2) Les vêtements
- 3) La maison
- 4) Les meubles de La maison
- 5) Les aliments et boissons des repas
- 6) Les objets placés sur la table et dont on se sert à tous les repas de la journée
- 7) La cuisine, ses meubles et les ustensiles qui s'y trouvent
- 8) L'école, ses meubles et son matériel scolaire
- 9) Le chauffage et l'éclairage
- 10) La ville
- 11) Le village ou le bourg
- 12) Les moyens de transport
- 13) Les travaux des champs et du jardinage
- 14) Les animaux
- 15) Les jeux et distractions
- 16) Les métiers

Pero la pregunta que surge inmediatamente es por qué sólo dieciséis centros de interés y por qué razón éstos específicamente. Los centros de interés pueden ser diferentes dependiendo de la edad de los informantes o del nivel cultural de los mismos. Y por otra parte como muy bien dice Mackey podemos englobar en centros de interés todos los intereses humanos. Pero, ¿cuántos centros de interés existen?. Existen centros de interés cuya delimitación semántica es muy clara, como podrían ser los animales, pero, la cocina ¿no es parte de la casa? ¿Qué diferencia sustancial implica para el informante la ville y la village?.

Mackey<sup>3</sup> en sus estudios de disponibilidad ha buscado una adaptación mayor del cuestionario a las características

---

<sup>3</sup>. W. F. Mackey, Op. Cit., p. 29.

sociológicas concretas de su población y añade seis centros de interés más a los propuestos por Gougenhém que son:

- 17) L'Eglise et la vie paroissiale
- 18) Les sports et l'equipement sportif
- 19) Les outils
- 20) L'automobile et ses parties
- 21) Actions courantes
- 22) Les affaires

Dimitrijevic<sup>4</sup> nos presenta sus razones para su selección de once centros de interés basada según él, en limitaciones de tiempo para realizar la encuesta y también basándose en una selección de centros que considera lo suficientemente selectivos para que los niños no se encuentren con campos semánticos muy extensos como podría ser un campo titulado vacaciones y que haría que los niños centraran su interés quizás en un área solamente, o por el contrario, en un campo tan específico que podría convertirse en demasiado técnico para los niños, como la construcción de un barco.

Esa experiencia la tuvimos nosotros con el centro de interés "materiales" en nuestra encuesta piloto. Los niños lo encontraban muy técnico y no eran capaces de mencionar más de dos o tres palabras acerca del mismo.

Los once centros de interés seleccionados por Dimitrijevic son:

- 1) Animals
- 2) Countryside
- 3) Town
- 4) Entertainment
- 5) Jobs and professions

---

4. N Dimitrijevic, Loc. cit., p.15.

- 6) Science
- 7) Means of transport
- 8) Politics
- 9) Parts of the house
- 10) Food and drink
- 11) Clothes

P. E. Njock<sup>5</sup> para su estudio del basaa y del francés en el Camerún, utiliza diez centros de interés que son:

- 1) Les parties du corps
- 2) La nourriture et les boissons
- 3) Les vêtements (pour hommes et pour femmes)
- 4) La maison (materiaux de constrution et meubles)
- 5) Le travail au champ
- 6) La cuisine (ce qu'on y voit)
- 7) La salle du classe (ce que tu vois et/ou utilises)
- 8) Les moyens de locomotion (ce qui te permet de te déplacer)
- 9) La musique et la danse (les genres)
- 10) Les jeux et divertissements

Azurmendi<sup>6</sup> en su investigación plantea su interés en un acercamiento mayor al gran grado de diversificación de la vida diaria en el mundo actual y basa su cuestionario en quince centros de interés que son:

- 1) Partes del cuerpo
- 2) Prendas de vestir y de calzar
- 3) Partes u objetos que hay en una vivienda (casa o piso)
- 4) Alimentos, bebidas
- 5) Relaciones familiares (madre, tío) y no familiares (alumno, amigo)
- 6) Ciudad o pueblo sus partes; lo que se ve y se hace en ellos
- 7) Los transportes: medios y vehículos que se utilizan
- 8) Enseñanza: clases de centros y sus partes: objetos y materiales utilizados, asignaturas, etc.
- 9) Trabajo: clases de trabajos, de oficios, de

<sup>5</sup>. P. E. Njock, tomado de la investigación de M. J. Azurmendi, Loc. Cit., p. 144.

<sup>6</sup>. Ibid., p. 237.



- profesiones
- 10) Juegos, diversiones, entretenimientos
  - 11) Deportes: prendas de vestir y objetos que se utilizan
  - 12) Acciones que normalmente realizamos todos los días
  - 13) La naturaleza, las vacaciones
  - 14) Agricultura y vegetales. Ganadería y animales. Pesca y peces
  - 15) Dinero, bancos, comercio.

R. Galisson<sup>7</sup> pretende sustituir los centros de interés por temas predilectos ya que no le interesa un inventario detallado del léxico, sino una búsqueda sistemática del léxico funcional, aquél que podría realmente utilizarse en un discurso aceptable. Plantea que los centros de interés pueden variar grandemente dependiendo de la edad, el nivel de escolaridad o el grado de cultura, el medio social o geográfico, el sexo y la nacionalidad de los informantes.

Si coincidiésemos con Galisson perderíamos un vocabulario muy frecuente en el habla coloquial o quizás de familiaridad entre los niños, que se refiere a las palabras tabú, que aparecen -y ampliamente- en las encuestas que hemos realizado.

Por otra parte se necesitaría un estudio previo al estudio lexicográfico para descubrir los temas de interés para cada comunidad.

En el estudio realizado por López Morales<sup>8</sup> en San Juan se utilizaron diez centros de interés:

---

7. R. Galisson, Lexicologie et enseignement des langues, París, Hachette, 1979.

8. H. López Morales, Dialectología y Sociolingüística. Temas Puertorriqueños, Madrid, Hipanova, 1979, p. 177.

- 1) Alimentos
- 2) Animales
- 3) La casa
- 4) La cocina
- 5) El cuerpo humano
- 6) Materiales
- 7) Muebles
- 8) Naturaleza
- 9) La ropa
- 10) Transporte

Los mismos centros seleccionados por López Morales son los que utiliza Román<sup>9</sup> en su investigación para propósitos comparativos.

Canizal Arévalo<sup>10</sup> limita a seis los centros de interés y como ya señaláramos, en el sexto el rango de coherencia semántica es muy alto, pues se limita a recoger adjetivos, los centros son:

- 1) Las partes del cuerpo
- 2) Objetos que hay en el interior de la casa
- 3) Alimentos y bebidas
- 4) Medios de transporte
- 5) Profesiones y oficios
- 6) Defectos físicos y morales (adjetivos)

Echeverría<sup>11</sup> hace una distinción entre sus centros de interés entre los de alto grado de coherencia semántica y bajo grado de coherencia semántica; los clasifica a priori. Estudia 10 centros: los primeros siete comparten la clasificación de alto grado y los restantes tres la de bajo grado. Estos son:

---

<sup>9</sup>. B. Román, Op. Cit.

<sup>10</sup>. A. V. Canizal Arévalo, Op. Cit..

<sup>11</sup>. M. Echeverría, Loc. Cit., p. 6.

- 1) Cuerpo humano
- 2) Medios de transporte
- 3) Ropa
- 4) Muebles
- 5) Herramientas y materiales de construcción
- 6) Alimentos
- 7) Animales
- 8) Naturaleza
- 9) La casa
- 10) La cocina

En el estudio realizado por Justo Hernández<sup>12</sup> encontramos un alto grado de coherencia semántica pues analiza la disponibilidad léxica de los hablantes dentro del campo semántico de los colores (adjetivos). Lleva a cabo su investigación en diez centros de interés, siempre desde la misma perspectiva:

- 1) Ropa
- 2) Zapatos
- 3) Cosmético
- 4) Automóviles
- 5) Anuncios comerciales
- 6) Mochilas, maletas, bolsas
- 7) Muebles
- 8) Tapices y alfombras
- 9) Pintura de casa (exterior e interior)
- 10) Vestido

3. Centros de interés en nuestra investigación. El objetivo de nuestra investigación no es el de escoger o recomendar unos centros de interés o una metodología para su determinación. Sin embargo, nos resultaba necesaria realizar una selección adecuada para obtener nuestro corpus.

Las metas que perseguimos al diseñar nuestros centros de interés fueron dos. Por un lado, queríamos tener una visión

---

12. H. G. Justo Hernández, Op. Cit.

abarcadora del universo léxico, donde diésemos la oportunidad a los informantes de responder el mayor número posible de unidades léxicas. Por otro lado, deseábamos evitar la repetición de entradas léxicas originadas por centros de interés que fueran percibidos por los informantes como muy próximos semánticamente.

Los puntos de referencia que tuvimos en cuenta para nuestra selección fueron tres: investigaciones anteriores; investigaciones realizadas en Puerto Rico; marco de referencia de nuestra muestra.

En relación con las investigaciones previas el Cuadro 8 nos resume los centros escogidos en los trabajos empleados como referencia.

Política	+		
Relaciones	+		
Trabajo	+		
Transporte			
		16	

1. J. Gougenhén et al.
2. W. P. Mackey
3. S. Dimitrijevic
4. P. E. Njock
5. S. J. Azumendi

Centros de interés escogidos por diferentes autores

A los efectos de comparación, los centros de interés fueron seleccionados con una sola palabra...

	1	2	3	4	5	6	7	8	9
acciones		+							
agricultura					+				
alimentos	+	+	+	+	+				
animales	+	+	+			+	+	+	+
barrio	+	+				+	+		+
campo	+	+	+						
casa	+	+	+	+	+	+	+	+	+
ciencia			+					+	+
ciudad					+				
cocina	+	+		+					
cuerpo humano	+	+		+	+	+			+
deportes		+			+	+	+	+	+
defectos					+				
energía	+	+						+	
entretenimiento	+	+	+	+	+				
escuela	+	+		+	+				
ganadería					+				
herramientas	+	+							
iglesia		+							
materiales									
mesa	+	+				+	+		+
música				+					
muebles	+	+							
naturaleza							+		
negocios					+	+	+		+
pesca					+				
política			+						
pueblo	+	+							
relaciones					+				
ropa	+	+	+	+	+	+	+		+
temas		+							
trabajos	+	+	+	+	+				
transporte		+	+	+	+	+	+	+	+
	16	22	10	10	15	10	10	6	10

- |                         |                          |
|-------------------------|--------------------------|
| 1. G. Gougenhéim et al. | 6. H. López Morales      |
| 2. W. F. Mackey         | 7. B. Román              |
| 3. N. Dimitrijevic      | 8. A. V. Canizal Arévalo |
| 4. P. E. Njock          | 9. M. Echeverría         |
| 5. M. J. Azurmendi      |                          |

Cuadro 8  
Centros de interés empleados  
por diferentes investigadores

A los efectos de comparar tratamos de describir los centros con una sola palabra que, desde nuestro punto de

vista, abarcara campos comunes en los diferentes trabajos y que no necesariamente coincidiera con la denominación original.

La última fila del cuadro nos muestra el número de centros empleados en las diferentes investigaciones. Llama la atención su variación que va, desde los seis centros empleados por Canizal Arévalo en su estudio del español de México, a los veintidós centros empleados por Mackey en su estudio comparativo del francés de Francia y de Canadá.

Las cruces en el cuadro marcan los centros que fueron utilizados por los diferentes autores. Resulta evidente que hay un grupo de ellos comunes; concretamente, de un total de treinta y tres centros hay once que se repiten y ocho de ellos fueron los que empleamos en nuestra encuesta piloto, en adición a "muebles" y "materiales". En la encuesta final mantuvimos siete de ellos.

Con respecto a las investigaciones llevadas a cabo en Puerto Rico, las dos referencias emplean los mismos centros de interés, que fueron propuestos por López Morales. Todos ellos formaron parte de nuestra encuesta piloto y en nuestra encuesta final mantuvimos siete de los diez por ellos utilizados.

Al comenzar nuestra investigación realizamos una encuesta piloto con un grupo de estudiantes de escuela elemental en la zona de Mayagüez. La encuesta se llevó a cabo en la escuela Cuesta de Piedra que es considerada de nivel socio-económico medio por el Departamento de Instrucción Pública.

Para dicha encuesta se utilizaron, como ya señaláramos, los mismos centros de interés que había utilizado López Morales en San Juan.

Las respuestas reflejaron que algunos centros no resultaban del todo claros para los estudiantes. Por ejemplo, en "materiales" no lograban identificar correctamente si nos referíamos a los materiales escolares que utilizaban como lápiz, acuarelas etc. o a materiales de construcción. Es probable que la mala interpretación se haya debido a que nuestra explicación se limitó a señalarles que escribieran los nombres de todos los materiales que conocieran. Los estudiantes que interpretaron el centro como materiales escolares escribieron mayor número de palabras que aquéllos que lo interpretaron como materiales de construcción. Por lo tanto, supusimos que éste último resultaba un tanto especializado para los niños.

Por otra parte la repetición de vocablos en los centros de interés "casa" y "muebles" nos hizo pensar que lo veían como una unidad y no podían hacer claramente la distinción entre el edificio y las cosas que en él había. Lo mismo ocurría con la "cocina", que la integraban como parte de la casa y las palabras volvían a repetirse, por ejemplo: si la lavadora y la secadora estaban en la cocina entraban en ese centro, pero también en el centro de interés "casa", y volvían a repetirse en el centro de interés "muebles".

Por tal motivo pensamos que un centro de interés "casa" era lo suficientemente abarcador como para no perder

entradas léxicas y al mismo tiempo ellos lo identificaban perfectamente tanto con la estructura, como con el mobiliario y todos los objetos que en ella podían encontrarse, muebles, electrodomésticos, etc..

Debido a que estamos realizando un estudio, independiente del presente, sobre disponibilidad léxica en el campo de los colores y también a nuestro interés en contrastar el trabajo de Justo Hernández<sup>13</sup>, sobre el campo "colores" en el español de México, es que incluimos el centro colores con el propósito de disponer de datos adicionales.

Para nuestra investigación final tomamos por tanto en consideración ocho centros de interés :

1. Alimentos
2. Animales
3. La casa
4. La ropa
5. Naturaleza
6. Colores
7. El cuerpo humano
8. Juegos y entretenimientos

4. Segmentación de la unidad léxica. En nuestro primer capítulo planteamos algunas observaciones sobre la dificultad para segmentar el término palabra.

Muller<sup>14</sup> nos expresa que no puede darse una definición satisfactoria del elemento palabra y por tanto una investigación estadística sobre el léxico se volvería

---

13. H. G. Justo Hernández, Op. Cit.

14. Ch. Muller, Estadística lingüística, Madrid, Gredos, 1967.



imposible. Pero como también señala este autor, en otros campos también resulta difícil definir otros conceptos como "adulto", "enfermedad" o "riqueza" y se realizan estudios basándose en convenciones precisas pero arbitrarias. Las normas pueden variar según los puntos de vista (pedagógico, lexicográfico, traducción automática etc.).

Entendemos que el criterio debe ser aceptar cualquier norma como válida si ésta resulta útil y razonable y se especifica desde un comienzo. Trabajamos, pues, basándonos en nuestra propia norma.

Usamos listas lematizadas y seguimos el criterio de respetar la entrada léxica que el informante ofreció. En los términos en que se presentó la opción de género como derivación y no como lexema diferente tomamos como unidad el masculino. Una, dos o más palabras se tomaron como una unidad simple así, por ejemplo, sofá se consideró una unidad y sofá-cama también. Las lexicalizaciones como en el caso de el-juez-le-dijo-al-cura, también se consideran unidad, de esta manera resulta válida la definición de palabra como el material gráfico entre dos espacios en blanco.

Al analizar parte de nuestros datos encontramos que aparecían, especialmente en los centros de interés "la ropa" y "alimentos", nombres propios que los estudiantes tomaban como comunes; consideramos a éstos también como comunes.

Nuestra investigación léxica, pues, tomará como clase única los sustantivos, que como expresáramos con anterioridad son las palabras más temáticas, con excepción

del centro de interés "colores" donde, por supuesto, sólo encontramos adjetivos. Al entrar los datos en el computador introdujimos guiones en los espacios para indicar que se consideran como una sola palabra.

5. Indice de cohesión. Echeverría<sup>15</sup> propone en su investigación la aplicación de un índice de cohesión de un centro para medir el grado de coincidencia en las respuestas del grupo de informantes para un mismo centro de interés. Este índice es una medida de dispersión que es la razón entre el rango promedio de los informantes y el rango del grupo para el centro:

$$IC(c) = \frac{\bar{r}}{R}$$

Siendo R el rango del grupo para el centro c y  $\bar{r}$  el promedio de los rangos de los individuos para dicho centro. En términos de las fórmulas presentadas en la Sección 2 del Capítulo III tendríamos:

$$IC(c) = \frac{D_{2g}(c)}{D_{1g}(c)}$$

Este tendría valor uno si todos los informantes coincidiesen y respondieran exactamente las mismas palabras, y valor más cercano a cero en caso de no existir coincidencia alguna. Cuanto más alto sea el índice mayor será el grado de coincidencia en las respuestas del grupo.

15. M. Echeverría, Loc. Cit., p. 26.

Aplicamos este índice a los ocho centros de interés para el grupo femenino del Colegio La Merced y los resultados se muestran en el Cuadro 9.

Nótese que el centro de interés "alimentos" (LFA) comparte con el centro de interés "juegos" (LFJ) el grado de dispersión más alto 0.17 y 0.16 respectivamente. En el centro "alimentos" (LFA) el número de entradas léxicas por informante es superior a todos los restantes.

El centro de interés "colores" (LFC), como era razonable suponer, presenta el mayor índice de cohesión pues tiene el menor número de entradas léxicas. En otras palabras se repiten las mismas entradas en casi todos los informantes. También en este centro hay un alto grado de coherencia semántica, ya que sólo obtuvimos adjetivos.

CENTRO	<u>r</u>	R	IC
LFA	16.0	91	0.17
LFN	19.2	83	0.23
LFL	18.2	75	0.24
LFR	14.6	49	0.29
LFT	16.6	80	0.20
LFC	14.5	32	0.45
LFH	17.3	67	0.25
LFJ	11.2	68	0.16

Cuadro 9  
Índice de cohesión para todos  
los centros del grupo femenino  
de la escuela "La Merced"

Queremos señalar que al analizar los datos encontramos una repetición muy alta de entradas léxicas en el centro de interés "naturaleza" (LFT), que ya se habían presentado en el centro de interés "animales" (LFN).

También aplicamos la fórmula del índice de cohesión al centro "alimentos" para todos los grupos femeninos de nuestra encuesta. El Cuadro 10 contiene los resultados obtenidos.

CENTRO	$\bar{x}$	R	IC
AFA	22.9	101	0.22
LFA	16.0	91	0.17
MFA	23.3	130	0.17
SFA	12.1	81	0.14
TFA	24.0	103	0.23

Cuadro 10  
Índice de cohesión para el centro  
"alimentos" de los grupos femeninos

En relación con este cuadro podemos apreciar que el rango (R) es muy superior en la escuela Modesto Rivera, clasificada como de nivel socioeconómico medio por el Departamento de Instrucción Pública.

Nótese también, comparando el Cuadro 10 con el Cuadro 9, que la variación es mucho mayor entre los centros de un mismo grupo, que va de 0.16 a 0.45, que la variación de un centro de interés para diferentes grupos en que ésta oscila de 0.14 a 0.23. Lo cual nos sugiere, como es lógico, que el índice de cohesión depende directamente del centro de interés o la forma en que el grupo lo interpreta y no del grupo en sus aspectos sociolingüísticos.

## CONCLUSIONES

Como planteáramos en la Introducción, nuestro objetivo al comenzar esta investigación era analizar tanto el aspecto teórico como el empírico de las fórmulas que se habían propuesto para calcular el índice de disponibilidad de las palabras.

Emprendimos el trabajo recopilando todas las investigaciones que estaban relacionadas con disponibilidad y descubrimos que por mucho tiempo sólo se había considerado la frecuencia de uso de las palabras para ordenarlas. Luego nos centramos en la nueva posibilidad de combinar la frecuencia con el orden de mención. Fueron varios los trabajos que estudiamos cuidadosamente porque empleaban el nuevo concepto de los índices de disponibilidad de las palabras. Al analizar el camino en ellos recorrido nos surgieron nuevas aplicaciones y posibilidades. A plantearlas dedicamos la mayor parte de nuestro esfuerzo.

Nuestras conclusiones podemos dividir las en aquéllas derivadas de la comparación de fórmulas ya existentes para el cálculo de índices de disponibilidad léxica; aquéllas acerca de las nuevas medidas que proponemos; las relacionadas con la metodología del procesamiento de datos y, por último, las sugerencias para investigaciones futuras.

Resultados de la comparación de fórmulas. Las fórmulas que sólo consideran la frecuencia absoluta provienen

básicamente de los trabajos que calificamos en el Capítulo I como pioneros en léxico-estadística. Encontramos que, por un lado, las ordenaciones inducidas por las fórmulas antes mencionadas y por otro, la basada en el rango medio, constituyen los polos opuestos entre las posibles ordenaciones.

Comparamos los resultados de calcular las fórmulas de López Morales y Lorán (L11 y L12) para listas abiertas y cerradas y llegamos a la conclusión de que el orden que se logra recortando las listas y aplicando LL1 es muy próximo al que se obtiene con LL2 si -por supuesto- se empleara la misma tasa de sustitución. Ambas fórmulas L11 y LL2 producen ordenaciones altamente correlacionadas cuando empleamos la fórmula de Spearman.

La fórmula LL2 puede causar algunas irregularidades en la ordenación cuando la tasa de sustitución es alta -por ejemplo 0.9-, y se producen descensos bruscos en el número de informantes que respondió en posiciones consecutivas. Según probamos, esto no ocurre con tasas de sustitución suficientemente bajas. Nuestra experiencia indica que 0.9 es un valor un tanto alto, lo cual coincide con lo expresado por López Chávez y Strassburger Frías, que proponen un valor mucho menor.

No encontramos grandes diferencias entre las fórmulas de López Morales-Lorán y la de López Chávez-Strassburger; aparte de que ambas toman en consideración la frecuencia y el orden de mención de las palabras, cuando las comparamos,

empleando la correlación de Spearman, confirmamos que la relación dependía mayormente del parámetro llamado tasa de sustitución y que si se emplea el valor próximo a 0.75 las fórmulas producen ordenaciones muy similares.

Sobre nuevos índices. Los trabajos pioneros de López Morales-Lorán y de López Chávez-Strassburger dejaron planteada una nueva línea de investigación dirigida a calcular un índice de disponibilidad para un informante. Estimulados por el pensamiento de que el tema podría resultar muy interesante principalmente para estudios sicolingüísticos, avanzamos en esa dirección.

Partimos de allí para proponer nuevos índices que midiesen la disponibilidad de un grupo, lo que creemos abre enormes posibilidades de hacer comparaciones sociolingüísticas. Al comienzo pensamos que para realizar comparaciones debíamos seleccionar las variables sociolingüísticas pertinentes y calcular el índice de disponibilidad de las palabras para los diferentes grupos. Luego de numerosas pruebas constatamos que los resultados no resultaban lógicos. Por lo tanto, llegamos a la conclusión de que no era posible separar los datos en relación con las variables seleccionadas, sino que había que calcular la disponibilidad de todas las palabras informadas y después relacionar las palabras con los grupos que correspondían a las variables sociolingüísticas.

Por otra parte pusimos en práctica algunas de las

fórmulas propuestas y sin lugar a dudas concluimos que las mismas pueden resultar de suma utilidad para inferencias sociolingüísticas como puede apreciarse en los cuadros ejemplificadores del Capítulo IV.

Como se puede apreciar a lo largo del Capítulo III existen diferentes índices posibles que reflejan la disponibilidad de un grupo para un centro de interés y cada uno toma en cuenta distintos aspectos, que deben considerarse y escogerse de antemano. A estos aspectos se deben las diferentes ordenaciones generadas. El investigador deberá decidir el índice más adecuado para su caso específico.

Sobre la metodología del procesamiento de datos.  
Consideramos altamente conveniente el uso de "hojas de cálculo electrónico", en lugar de desarrollar programación específica, debido a la gran flexibilidad en la realización de los cálculos y en la producción de los listados. Obviamente en función del volumen de datos es importante contar con un equipo con las capacidades adecuadas.

En la manipulación de grandes cantidades de datos y con un corpus dividido en varios subconjuntos como el nuestro, elegir y mantener una notación clara, concisa y constante, tanto en el trabajo manual como en la computadora, resulta absolutamente imprescindible.

Cuando se van a calcular los diferentes índices para las diferentes matrices de frecuencia nos resultó lo más



conveniente mantener las matrices en hojas de cálculo separadas, pero, con idéntico formato. De forma tal que pudiésemos utilizarlas como insumo en otra hoja de cálculo electrónica que sólo contuviese fórmulas; así éstas últimas al combinarse adoptan automáticamente los valores correspondientes en cada caso. Esta conclusión es fruto de innumerables horas de trabajo mal utilizadas y esperamos que pueda servir de ayuda a otros investigadores en el futuro.

También es muy importante en la metodología del procesamiento de los datos analizar, antes de comenzar con los cómputos, qué cálculos son pre-requisitos de cuáles, y qué orden es el más conveniente a efectos de evitar la duplicación del trabajo y minimizar la posibilidad de errores.

Investigaciones futuras. Al ser nuestra investigación más teórica que empírica -como es lógico- se nos plantean muchas y variadas posibilidades para comprobar y aplicar los índices propuestos.

En particular, creemos que el cálculo de índices de disponibilidad léxica para un grupo es un camino promisorio y es nuestro propósito comenzar a caminar en ese sentido.

A la vez que nos preocupan los resultados obtenidos con la comparación de la disponibilidad léxica de niños del interior de la isla y de la zona metropolitana, consideramos que puede ser relevante con propósitos pedagógicos investigar con una muestra apropiada si se confirman las

observaciones expuestas en el Capítulo IV.

Por otra parte, los estudios semánticos en relación con los índices de disponibilidad léxica son hoy por hoy, prácticamente desconocidos; el índice de cohesión propuesto por Echeverría abre buenas posibilidades de investigación en el campo semántico. Una investigación sobre los centros de interés más adecuados a nuestro contexto cultural, y tomando en cuenta los campos referenciales que no se han estudiado en las investigaciones anteriores, podría resultar de enorme interés para viabilizar estudios futuros.

APENDICE A

APLICACION DE DIVERSAS FORMULAS

GRUPOS DE SEXO MASCULINO Y CENTRO DE INTERES ALIMENTOS



LASAGNA	1	43	0.036792	46	0.002153	44	0.010268	46	0.036792	54	0.034391	50	8.00	33
LECHE	3	17	0.183469	11	0.103508	9	0.133714	9	0.183469	18	0.179962	14	3.33	6
LECHON	3	19	0.099707	22	0.004221	38	0.023745	36	0.099707	28	0.092372	25	9.00	39
LECHUGA	1	49	0.033113	50	0.001292	50	0.007701	52	0.033113	58	0.030655	55	9.00	40
MACARRONES	1	55	0.000000	63	0.000067	69	0.001909	69	0.029413	61	0.013706	68	16.00	66
MATZ	3	18	0.000000	64	0.000756	53	0.009377	49	0.073954	35	0.062774	35	12.33	57
MANGO	2	26	0.069231	30	0.046352	16	0.060572	21	0.094906	29	0.087919	26	7.50	24
MANTECADO	1	63	0.069231	31	0.046154	17	0.057692	23	0.069231	36	0.068567	33	2.00	2
MANTEQUILLA	1	65	0.062308	33	0.027692	22	0.043269	27	0.062308	38	0.061118	37	3.00	4
MANZANA	9	3	0.414631	3	0.186035	4	0.283541	2	0.462909	3	0.448658	2	5.44	16
MELOCOTON	2	27	0.026821	55	0.000744	54	0.007581	54	0.050961	42	0.046067	42	11.50	54
MELON	3	22	0.092109	23	0.028535	20	0.050954	25	0.121522	22	0.102149	23	9.67	43
MUSLO	2	30	0.056077	36	0.016686	27	0.034958	30	0.102402	27	0.066695	34	10.50	46
PAN	1	57	0.000000	65	0.000198	64	0.004387	65	0.025675	68	0.019352	65	13.00	62
PAPAS-FRITAS	2	36	0.000000	66	0.000217	62	0.004387	57	0.051704	41	0.032626	52	14.50	64
PAPAS-MAJADRS	1	51	0.076923	29	0.076923	12	0.076923	15	0.076923	34	0.076923	29	1.00	1
PAPAYA	2	32	0.000000	67	0.000477	55	0.006129	55	0.049815	44	0.041063	44	12.50	59
PAPA	5	13	0.166862	13	0.023748	23	0.066339	17	0.195458	16	0.173484	15	8.60	36
PAVO	1	41	0.000000	68	0.000279	61	0.003249	63	0.024139	70	0.021711	63	12.00	56
PEPINILLO	1	39	0.026821	56	0.000465	58	0.004332	60	0.026821	65	0.024357	60	11.00	52
PERA	6	10	0.236631	8	0.064393	13	0.123064	10	0.265227	10	0.242613	9	7.50	25
PERRO-CALIENTE	4	16	0.249342	7	0.154358	5	0.191575	7	0.249342	11	0.245190	8	3.25	5
PINCHO	1	68	0.000000	69	0.000042	71	0.001879	71	0.041693	48	0.010890	71	18.00	70
PIZA	2	29	0.080271	26	0.010744	31	0.030115	32	0.080271	32	0.075885	30	7.50	26
PLATANO	1	38	0.056077	37	0.016615	28	0.032452	31	0.056077	40	0.054478	39	4.00	9
POLLO	2	33	0.118385	19	0.044308	18	0.075721	16	0.118385	23	0.115596	19	3.50	7
QUESO	7	6	0.356808	5	0.103909	8	0.193792	6	0.356808	8	0.344437	6	5.14	15
QUENEPA	3	23	0.045422	40	0.006171	36	0.022920	37	0.114856	24	0.072752	32	12.33	58
REMOLACHA	3	20	0.113384	20	0.012036	30	0.037816	29	0.113384	25	0.106541	20	8.00	34
SOPA	1	58	0.000000	70	0.000071	68	0.002506	66	0.046326	45	0.012217	69	17.00	67
TAMARINDO	1	70	0.026821	57	0.000465	59	0.004332	61	0.026821	66	0.024357	61	11.00	53
TOMATE	1	40	0.036792	47	0.002153	45	0.010268	47	0.036792	55	0.034391	51	8.00	35
TORONJA	6	7	0.126759	17	0.021363	25	0.058867	22	0.204955	15	0.170694	17	10.67	47
TOSTADA	1	52	0.033113	51	0.001292	51	0.007701	53	0.033113	59	0.030655	56	9.00	41
TOSTADA	7	56	0.000000	71	0.000123	65	0.009770	48	0.391524	6	0.049404	40	22.00	71
UVA	5	14	0.168517	12	0.017984	26	0.063716	19	0.192656	17	0.181255	13	7.80	28
ZANAHORIA	1	59	0.029802	52	0.000775	52	0.005776	56	0.029802	60	0.027325	57	10.00	45











QUESO	2	39	0.004693	55	0.000462	52	0.006363	50	0.060604	42	0.051157	43	14.00	47
RAVIOLI	1	84	0.019775	35	0.005184	34	0.019775	35	0.059049	44	0.055731	39	7.00	8
REPOLLO	2	40	0.062500	24	0.040035	16	0.063753	22	0.113163	27	0.105928	28	10.00	26
SALCHICHA	1	85	0.019775	36	0.005184	35	0.019775	36	0.059049	45	0.055731	40	7.00	9
SALSA	1	86	0.083333	15	0.066667	11	0.083333	15	0.100000	32	0.099041	29	2.00	2
SANDWICH	2	41	0.004693	56	0.000462	53	0.006363	51	0.060604	43	0.051157	44	14.00	48
SOPA	1	87	0.011124	43	0.001866	42	0.011124	44	0.047830	54	0.044280	52	9.00	22
TACO	1	88	0.035156	32	0.014400	29	0.035156	32	0.072900	39	0.070143	38	5.00	4
TAMARINDO	2	42	0.008343	46	0.001121	45	0.008543	45	0.059000	46	0.046502	47	17.50	66
TOCINETA	1	89	0.000000	90	0.000087	68	0.001980	68	0.025419	73	0.022210	70	15.00	55
TOMATE	5	9	0.175293	3	0.083008	7	0.175293	4	0.354708	4	0.340957	4	5.40	5
TORONJA	1	90	0.000000	91	0.000015	79	0.000805	79	0.021442	82	0.014021	79	19.00	72
TOSTON	1	91	0.014832	38	0.003110	37	0.014832	38	0.053144	50	0.049676	45	8.00	11
UVA	3	28	0.039242	29	0.008170	33	0.039242	29	0.149925	21	0.139480	21	8.67	15
VERDURA	2	43	0.038676	30	0.014642	28	0.038676	30	0.104281	31	0.098096	30	9.00	23
YAUTIA	3	29	0.045834	28	0.011626	30	0.045834	28	0.156486	18	0.146272	19	8.33	12
YUCA	2	44	0.173611	5	0.151111	3	0.173611	5	0.201111	14	0.199393	11	2.00	3
ZANAHORIA	6	4	0.175293	4	0.083067	6	0.176963	3	0.380444	3	0.360754	3	7.17	10



1	80	0.000000	64	0.000259	66	0.003017	67	0.022415	78	0.018816	72	12.00	43
5	19	0.071094	12	0.051372	15	0.086570	14	0.178406	14	0.164435	14	8.60	26
1	81	0.050000	19	0.040000	20	0.050000	21	0.060000	36	0.059424	33	2.00	2
1	82	0.000000	65	0.000001	105	0.000134	105	0.008863	105	0.004734	105	24.00	105
1	83	0.000000	66	0.000720	59	0.005363	60	0.027673	68	0.023682	66	10.00	31
1	84	0.000000	67	0.000036	84	0.001028	86	0.015838	92	0.011878	91	16.00	69
2	47	0.000000	68	0.000006	93	0.000531	91	0.021105	82	0.012809	87	21.50	96
1	85	0.021094	29	0.008640	33	0.021094	33	0.043740	50	0.042086	43	5.00	6
2	48	0.000000	69	0.000039	81	0.001292	81	0.025969	70	0.018562	76	18.50	83
1	86	0.000000	70	0.003110	46	0.011865	48	0.035429	60	0.033438	58	7.00	19
4	22	0.021094	30	0.010096	30	0.032335	30	0.114214	23	0.100375	22	10.25	32
2	49	0.000000	71	0.000082	76	0.002142	74	0.031852	63	0.023914	65	16.00	70
5	20	0.031641	27	0.015538	26	0.054411	19	0.165923	16	0.151597	15	8.40	24
1	87	0.000000	72	0.000060	78	0.001371	79	0.017598	89	0.013326	85	15.00	61
14	2	0.361914	2	0.264675	2	0.381686	2	0.609741	2	0.580666	2	6.21	13
1	88	0.000000	73	0.000155	72	0.002263	73	0.020174	83	0.016772	81	13.00	51
3	31	0.000000	74	0.000757	55	0.006525	56	0.052374	43	0.040294	46	16.67	77
1	89	0.000000	75	0.001200	52	0.007151	52	0.030748	64	0.026568	63	9.00	27
2	50	0.000000	76	0.000025	85	0.001035	83	0.024385	74	0.017272	78	19.00	86
6	15	0.000000	77	0.002626	47	0.019223	37	0.127500	20	0.102236	21	13.67	55
11	5	0.000000	78	0.001805	51	0.016424	38	0.168112	15	0.124639	18	17.73	80
2	51	0.000000	79	0.000021	88	0.001012	87	0.024374	76	0.017850	77	18.50	84
6	16	0.050000	20	0.040606	18	0.057932	18	0.143747	19	0.122981	19	14.00	59
2	52	0.000000	80	0.000101	73	0.001961	75	0.028416	67	0.019684	71	19.00	87
6	17	0.114844	9	0.065750	13	0.133860	9	0.261117	9	0.249490	8	5.33	9
12	3	0.214844	5	0.144720	5	0.237940	4	0.452446	3	0.421902	3	8.08	23
3	32	0.000000	81	0.000185	70	0.003467	65	0.045973	49	0.035772	52	16.33	75
1	90	0.000000	82	0.000001	104	0.000178	104	0.009848	104	0.005311	104	23.00	104
1	91	0.000000	83	0.002000	50	0.009535	51	0.034164	61	0.029806	60	8.00	22
3	33	0.037500	22	0.024008	21	0.038034	24	0.073522	30	0.065601	30	15.67	65
7	12	0.112500	10	0.075644	9	0.130349	10	0.250751	12	0.233138	11	9.29	29
3	34	0.021094	31	0.012182	29	0.036981	29	0.104075	25	0.096633	24	7.67	21
2	53	0.000000	84	0.000044	80	0.001462	78	0.027384	69	0.020291	70	17.50	78
1	92	0.000000	85	0.000002	100	0.000238	100	0.010942	98	0.005958	100	22.00	99
9	7	0.078125	11	0.058841	14	0.106572	11	0.264123	8	0.235230	10	11.44	40
1	93	0.037500	23	0.024000	23	0.037500	27	0.054000	39	0.052969	35	3.00	3
3	35	0.000000	86	0.003631	39	0.020708	34	0.089817	27	0.077483	28	9.33	30
1	94	0.000000	87	0.000259	67	0.003017	68	0.022415	79	0.018816	73	12.00	44
1	95	0.000000	88	0.000001	106	0.000134	106	0.008863	106	0.004734	106	24.00	106

MONDONGO	2	54	0.015820	36	0.005616	34	0.019843	36	0.064272	32	0.058623	34	8.50	25
MORCILLA	3	36	0.066667	13	0.066669	12	0.067005	17	0.085585	29	0.076844	29	16.00	71
MUSLO	1	96	0.000000	89	0.000060	79	0.001371	80	0.017598	90	0.013326	86	15.00	62
PANA	1	97	0.000000	90	0.000003	98	0.000264	98	0.010131	100	0.006684	98	21.00	94
PANQUEQUE	2	55	0.000000	91	0.002021	48	0.010306	49	0.048418	47	0.040394	45	12.50	48
PAN	7	13	0.015820	37	0.005324	35	0.020413	35	0.123794	21	0.099522	23	15.71	66
PAPA	3	37	0.000000	92	0.003128	41	0.012796	43	0.059515	37	0.050375	37	15.00	63
PARCHA	1	98	0.000000	93	0.001200	53	0.007151	53	0.030748	65	0.026568	64	9.00	28
PASTELILLOS	2	56	0.000000	94	0.000267	65	0.003450	66	0.033961	62	0.027228	62	15.50	64
PATITAS-DE-CERDO	2	57	0.037500	24	0.024003	22	0.037764	25	0.064131	33	0.059653	31	12.00	45
PAVO	1	99	0.000000	95	0.000008	92	0.000434	94	0.011546	95	0.008412	95	19.00	88
PERA	9	8	0.015820	38	0.009686	32	0.046780	22	0.223235	13	0.189845	13	11.78	41
PESCAO	7	14	0.253320	3	0.230385	3	0.260605	3	0.332977	7	0.321784	7	6.43	14
PINCHO	1	100	0.000000	96	0.000013	89	0.000578	89	0.012829	93	0.009438	93	18.00	82
PINA	4	23	0.021094	32	0.009847	31	0.028835	31	0.096687	26	0.082110	27	14.00	60
PIZZA	4	24	0.066667	14	0.066924	11	0.070935	16	0.116241	22	0.103699	20	12.75	49
PLATANO	4	25	0.000000	97	0.000294	64	0.005356	61	0.066506	31	0.049558	38	16.25	73
POLLO	12	4	0.138737	7	0.102909	7	0.159572	7	0.362985	4	0.329260	5	10.42	34
POSTRE	2	58	0.000000	98	0.000025	86	0.001035	84	0.024385	75	0.017272	79	19.00	89
PULPO	2	59	0.043945	21	0.019584	25	0.043945	23	0.087966	28	0.084728	26	5.00	7
QUESO	3	38	0.000000	99	0.000757	54	0.006570	55	0.053358	41	0.040871	44	16.33	76
REMOLACHA	1	101	0.000000	100	0.000259	68	0.003017	69	0.022415	80	0.018816	74	12.00	46
SALCHICHA	5	21	0.066667	15	0.071818	10	0.089447	13	0.163355	17	0.149288	16	10.40	33
SAL	1	102	0.000000	101	0.000001	107	0.000134	107	0.008863	107	0.004734	107	24.00	107
SANDWICH	2	60	0.000000	102	0.000742	57	0.006134	58	0.041927	51	0.034270	53	13.50	53
SARDINA	1	103	0.000000	103	0.000002	101	0.000238	101	0.010942	99	0.005958	101	22.00	100
SOPA	1	104	0.000000	104	0.000259	69	0.003017	70	0.022415	81	0.018816	75	12.00	47
TAMARINDO	2	61	0.000000	105	0.000722	58	0.005601	59	0.038615	55	0.029640	61	16.00	72
TEMBLEQUE	1	105	0.037500	25	0.024000	24	0.037500	28	0.054000	40	0.052969	36	3.00	4
TOCINO	1	106	0.000000	106	0.000003	99	0.000264	99	0.010131	101	0.006684	99	21.00	95
TOMATE	4	26	0.036914	26	0.013838	28	0.037626	26	0.104798	24	0.093771	25	13.25	52
TORONJA	1	107	0.000000	107	0.000432	62	0.004022	63	0.024906	72	0.021109	68	11.00	37
TOSTON	2	62	0.000000	108	0.000022	87	0.000905	88	0.023117	77	0.015322	82	20.50	93
UVA	9	9	0.065625	17	0.044449	17	0.096386	12	0.255745	11	0.228111	12	11.78	42
YUCA	1	108	0.000000	109	0.000432	63	0.004022	64	0.024906	73	0.021109	69	11.00	38
ZANAHORIA	1	109	0.028125	28	0.014400	27	0.028125	32	0.048600	46	0.047215	41	4.00	5









APENDICE B

CORRELACION ENTRE DIVERSOS INDICES

GRUPOS DE SEXO MASCULINO Y CENTRO DE INTERES ALIMENTOS





```

---TMR---
FR  LL1-0.75  LL2-0.60  LL2-0.75  LL2-0.90  LS  NJ
1.0000000  0.6659978  0.5946170  0.6302437  0.7799172  0.7429527  0.2672878
LL1-0.75  1.0000000  0.9613473  0.9647715  0.9259118  0.9400701  0.8103838
LL2-0.60  1.0000000  0.9952540  0.9419334  0.9596751  0.8547221
LL2-0.75  1.0000000  0.9605192  0.9765249  0.8471731
LL2-0.90  1.0000000  0.9946329  0.7306737
LS  1.0000000  0.7745342
NJ  1.0000000

```





Escuela

1.  
2.  
3.  
4.  
5.  
6.  
7.  
8.  
9.  
10.  
11.  
12.  
13.  
14.  
15.  
16.  
17.  
18.  
19.  
20.  
21.  
22.  
23.  
24.  
25.  
26.  
27.  
28.  
29.  
30.  
31.  
32.  
33.  
34.  
35.  
36.  
37.  
38.  
39.  
40.  
41.  
42.  
43.  
44.  
45.  
46.  
47.  
48.  
49.  
50.

APENDICE C

FORMULARIO EMPLEADO EN LA RECOGIDA DE DATOS

Escuela : \_\_\_\_\_

F \_\_\_\_\_ M \_\_\_\_\_

	Alimentos	Animales	La casa	Juegos y entretenimientos
1.	_____	_____	_____	_____
2.	_____	_____	_____	_____
3.	_____	_____	_____	_____
4.	_____	_____	_____	_____
5.	_____	_____	_____	_____
6.	_____	_____	_____	_____
7.	_____	_____	_____	_____
8.	_____	_____	_____	_____
9.	_____	_____	_____	_____
10.	_____	_____	_____	_____
11.	_____	_____	_____	_____
12.	_____	_____	_____	_____
13.	_____	_____	_____	_____
14.	_____	_____	_____	_____
15.	_____	_____	_____	_____
16.	_____	_____	_____	_____
17.	_____	_____	_____	_____
18.	_____	_____	_____	_____
19.	_____	_____	_____	_____
20.	_____	_____	_____	_____
21.	_____	_____	_____	_____
22.	_____	_____	_____	_____
23.	_____	_____	_____	_____
24.	_____	_____	_____	_____
25.	_____	_____	_____	_____



La ropa

Naturaleza

Colores

El cuerpo humano

1.

2.

3.

4.

5.

6.

7.

8.

9.

10.

11.

12.

13.

14.

15.

16.

17.

18.

19.

20.

21.

22.

23.

24.

25.

APENDICE D

EJEMPLO DE MATRICES DE FRECUENCIAS  
ESCUELA "AMALIA EXPOSITO"















Alimentos

carne

carne de res

carne de cerdo

chuleta

carne

carne de res

carne de cerdo

carne de res

carne

carne

carne

carne

carne

carne

carne de res

carne

carne

carne de res

carne de cerdo

carne de res

carne de cerdo

APENDICE E

FORMULARIOS DE CENTROS DE INTERES

GRUPO DE SEXO MASCULINO DE LA ESCUELA "LA MERCED"

(cinco estudiantes)

AlimentosAnimalesLa casaLa ropa

viste

emparedado

perro caliente

chuleta

arroz

avichuela

gambas

arroz chino

pollo

pina

hina

leche

carne

queso

papa

melon

patata

lechuga

patatas

vaca

cabra

caraballo

tortugas

aves

aguila

pollito

zorro

gallina

ganado

jirafa

cocodrilo

dinosaurio

papatamos

monjes

orangutan

pato

seal

peccari

nevera

cama

estufa

lavamanos

pasa manos

mesa

silla

escritorio

ventana

cama

televisión

plato

tasa

flores

espejo

garaje

patio

sala

antenna

traje

medias

zapatos

minifalda

ropa

guata

camisa

sudadera

tenis

chaleco

ropa

suite

choreo

pantalón

mao

sambor

corbata

sanda

Naturaleza

Colores

El cuerpo humano

Juegos y entretenimientos

árbol  
 cama  
 casta  
 casa  
 montaña  
 mesita  
 mango  
 lianas  
 pluma  
 flor  
 la vaca  
 familia  
 Hamburgo  
 plantas  
 Tomates  
 animales

rojo  
 azul  
 verde  
 amarillo  
 violeta  
 morado  
 oro  
 plata  
 gris  
 negro  
 blanco  
 naranja  
 negro  
 rosa  
 color carne

corazón  
 ombligo  
 codos  
 hombros  
 brazos  
 dedos  
 manos  
 ojos  
 pestañas  
 pelo  
 oídos  
 boca  
 nariz  
 pulmón  
 amígdalas  
 garganta  
 lengua  
 tráquea  
 estómago  
 intestinos  
 piel  
 tendones

juego pelota  
 bola  
 canicas  
 charita  
 puyo y  
 paracaídas  
 cubo  
 cocacola  
 cimen de  
 pisco  
 manojito  
 biblia

Nombre: \_\_\_\_\_

Sexo Masculino

Alimentos

Animales

La casa

La ropa

manzana

queso

pera

Pizza

embrocados

arroz

espaguetis

acarina

abichuelas

tomates

verduras

maiz

cafe

chuletas

carne

pancitas

panquecas

panes

huevos

pollitos

gallinas

perros

elefantes

leones

leopardos

conejos

Ratones

gusanos

Caballos

monos

sebras

jirafas

ardillas

colibris

golondrinos

Ajuna California

mariposas

celas

video

televisor

camara

camion

cañuter

estufa

mesa

taquias

platos

vañeras

cartinas

botiquin

sota

archivo

escritorio

silla

libros

si llon

orno

galaje

gramos

patio

antesala

visita

resibir

faldas

sudadera

zapatos

camiseta

medias

blusa

sofa

abrigo

chaleco

correa

agradador

gorras

sonbiero

pañuelo

carbante

Pantalón

Naturaleza

Colores

El cuerpo humano

Juegos y entretenimientos

lirinas

rojo

orejas

manipuli

ierba

gris

velis

pelotas

Montañas

negro

voca

balones

arboles

Verde

dientes

damos

ambes

Braun

corazon

seritos

ibos

anarajento

bencas

rescate

enes

asul

arterias

Correl

roquidien

violeta

dedos

INDIA

osa

rosita

maneceras

asociacion

agner

amarillo

tabillos

siclismo

ros

blanco

palmas de la

lagos

ti la

pies

mar

oro

rodillas

res

planta

espinas dorsales

baronias

transparente

espartera

partes

ojos

allenas

tinpuno

stones

seleto

reales

antos

artijos

es

Nombre: \_\_\_\_\_

Sexo masculino

Alimentos

Animales

La casa

La ropa

arroz

china

arroz con dulce

tomate

vinco

guineja

arroz

arroz

arroz

arroz

arroz de fuerza

vaca

leopardo

ratas

leon

elefante

conejo

gallina

hormiga

arroz trus

arroz cocido

arroz

arroz

arroz

arroz

arroz

cama

Baño

cocina

lampara

sofa

televisor

bañon

radio

mesa

estufa

videocase

la de casa

garage

patio

portones

antecala

Zapato

camisa

media

chaleco

bufanda

falda

sneter

pantalón

corbata

gasa

dota

mameluco

casaca

NaturalezaColoresEl cuerpo humanoJuegos y entretenimiento

rojo

rojo

corazon

pelota

verde

verde

pecho

bolon e

amarillo

amarillo

pulman

monos

blanco

Blanco

pie

tenis

violeta

violeta

ojo

simo

naranja

naranja

cadera

ant las fi's

azul

azul

dedo

recafe

plateado

plateado

Brazo

canta

rosa

rosita

radica

sedoble

cafe

cafe

castilla

ralo

naranja

cafe

nariz

sclisme

naranja

nariz

rojo

oreja

cintura

diente

sejas

pecho

mulo

Nombre: \_\_\_\_\_

Sexo Masculino

Alimentos

Animales

La casa

La ropa

papas rayadas  
 vino  
 manzana  
 arroz  
 pollo  
 pisco  
 chicha  
 uva  
 manzana  
 queso  
 papa  
 manzana  
 tomate  
 papas fritas  
 papa

avestruz  
 aguila  
 pollito  
 gallo  
 gallina  
 gato  
 perro  
 elefante  
 serpiente  
 cobra  
 paloma  
 zebra  
 caballo  
 imopotamo  
 tigre  
 leon  
 mono  
 delfin  
 cotorra  
 siguena

plancha  
 alfombra  
 silla  
 mueble  
 asiento  
 estufa  
 armario  
 escritorio  
 laborera  
 secadora  
 cama  
 vidriocasa  
 caset  
 comedor  
 nevera  
 ventana  
 garaje  
 telefono  
 antepala  
 puerta  
 closet

zapatos  
 camisa  
 chaqueta  
 sudadera  
 meter  
 media  
 correa  
 ropa interior  
 ropa exterior  
 gorra  
 sombrero  
 corbata  
 lirantes  
 maones  
 parateleros



Naturaleza

Colores

El cuerpo humano

Juegos y entretenimie

árbol  
 clavel  
 flor  
 rosa  
 clavel  
 mala madre  
 cielo  
 alba  
 calmas  
 montaña  
 cerros  
 valles  
 ríos  
 manifiestas  
 diluvios  
 mariposas

blanco  
 negro  
 gris  
 chinita  
 amarilla  
 verde  
 marón  
 carne  
 plateado  
 oro  
 azul  
 rosita  
 azul claro  
 transparente

corazón  
 cabeza  
 brazo  
 antebrazo  
 pierna  
 cintura  
 costilla  
 hueso  
 oreja  
 nariz  
 pestaña  
 ojo  
 pelo  
 pulmones  
 mente  
 dedos  
 uña  
 diente  
 lengua

monopolio  
 juego de  
 billar  
 truco  
 pelota  
 fútbol  
 baloncesto  
 rescate  
 en la calle  
 balones  
 manijeros  
 chabito  
 congelan  
 Hoki

Nombre: \_\_\_\_\_

Sexo masculino

Alimentos

café caliente

emparedado

melón

arroz

habichuela

patatas

pollo

guineas

carne de caballo

lechón

gandules

manzana

café

carne de res

papas

tomates

hamburguesas

Animales

león

tigre

caballo

yesca

zebra

jirafa

hipopótamo

rinoceronte

burro

toro

cabra

conejo

serpiente

ardilla

ratones

arce

mono

ganso

patos

gallos

risnes

La casa

closet

alarma

ventanas

puertas

paredes

hornos

muebles

gabinets

gabinetes

estufa

mueble

espejo

testadora

licuadora

lavadora

ceradora

chirero

botiquin

plancha

luzes

gabinets

masa

silla

portenas

masa

masa

La ropa

gabán

correa

guayabera

ropa interior

corbata

chaleco

camisa

pantalón

zapatos

abrigos

medias

gorra

visera

sombrero

lazos

corbatas

trajes

maillots

pantalones

Naturaleza

Colores

←

-El cuerpo humano

Juegos y entretenimiento

esto  
 doles  
 ato  
 nears  
 au lara  
 lo  
 shras  
 ner dider  
 ar  
 che  
 rufas  
 ges  
 der  
 uia  
 ee  
 ve va  
 ara ake  
 tidis  
 ipir ur  
 nteno

Verde  
 Azul  
 Rosita  
 Cicleta  
 Blanco  
 Rojo  
 Marro  
 Cafe  
 durre  
 Amarelo  
 Negro  
 Ori  
 chini te  
 dris tel

Piel  
 hirtetino  
 Mano  
 Pienas  
 Pie  
 Pelo  
 Ojos  
 adomen  
 zead  
 rypas  
 Cerazon  
 Organo  
 celan u vertedel  
 dico  
 Glnide las  
 gelgante  
 Mente  
 do  
 u na  
 Nazir

choly  
 Pimpel  
 Escuder  
 Barista  
 Pelota  
 Balon Pie  
 kite  
 Rescate  
 la tabla  
 Pesomaw  
 Brinco la ma  
 tira y tapet  
 nichismo

Handwritten text on the left margin, possibly a page number or reference.

Vertical column of handwritten text on the left margin, including names and possibly dates.

APENDICE F

DISPONIBILIDAD DE INFORMANTES  
APLICACION DE TRES DIFERENTES FORMULAS  
PARA LOS DIFERENTES CENTROS DE INTERES

---LAF---

Informante	Rango		dc(i) * LL2-0.75		dc(i) * LS	
1	16	11	1307	14	2533	14
2	20	3	2505	3	4230	2
3	15	12	1985	10	3440	10
4	7	15	996	15	1696	15
5	9	14	1655	12	2704	12
6	12	13	2033	9	3321	11
7	19	5	2199	6	4028	4
8	17	10	2254	5	3670	8
9	19	7	2115	8	3736	6
10	22	2	2782	1	4871	1
11	18	8	1972	11	3516	9
12	22	1	2524	2	4080	3
13	19	6	2127	7	3677	7
14	18	9	1436	13	2585	13
15	20	4	2324	4	4016	5

\* X 10,000

---LNF---

Informante	Rango		dc(i) * LL2-0.75		dc(i) * LS	
1	25	1	2901	5	5916	3
2	21	6	1663	14	3230	14
3	21	8	2336	11	4801	10
4	13	13	1447	15	2987	15
5	13	15	3201	3	4393	12
6	13	14	2664	7	4235	13
7	21	7	6531	1	7471	1
8	21	5	4929	2	7454	2
9	17	12	2255	12	4981	9
10	20	9	1962	13	4621	11
11	25	2	2929	4	5539	4
12	22	3	2623	8	5230	6
13	19	10	2753	6	5091	7
14	22	4	2548	10	5064	8
15	18	11	2618	9	5321	5

\* X 10,000

---LLF---

Informante	Rango		dc(i) * LL2-0.75	dc(i) * LS	
1	16	12	1469	13	3154
2	21	2	1829	9	3980
3	20	3	2563	2	4900
4	19	5	2753	1	4495
5	17	10	2402	4	4876
6	15	14	2386	5	4436
7	18	7	2426	3	4856
8	18	6	1714	11	3526
9	24	1	1469	14	3452
10	18	9	1826	10	3892
11	13	15	2036	7	3917
12	15	13	1990	8	3746
13	16	11	1299	15	2933
14	20	4	2385	6	4644
15	18	8	1518	12	3972

\* X 10,000

---LJF---

Informante	Rango		dc(i) * LL2-0.75		dc(i) * LS	
1	12	7	1767	8	2458	8
2	14	4	1363	14	2121	10
3	6	14	1588	10	1934	14
4	8	12	1580	11	2057	11
5	5	15	934	15	1416	15
6	8	13	1591	9	2021	13
7	16	2	1798	7	2614	4
8	10	9	1947	4	2575	7
9	9	10	1922	6	2447	9
10	12	6	1933	5	2613	6
11	9	11	2120	2	2747	3
12	11	8	1546	12	2043	12
13	16	1	2639	1	3993	1
14	13	5	2116	3	2820	2
15	15	3	1416	13	2614	5

\* X 10,000



---LRF---

Informante	Rango		dc(i) *		dc(i) *	
			LL2-0.75		LS	
1	13	13	1319	15	4096	13
2	17	2	2884	10	4667	5
3	15	6	3026	7	4658	6
4	11	15	2542	13	3756	15
5	14	9	3051	5	4731	3
6	14	8	3160	2	4968	2
7	12	14	2808	11	4364	11
8	15	5	3042	6	4519	8
9	18	1	3191	1	5006	1
10	14	10	2976	8	4474	9
11	16	4	3055	4	4684	4
12	13	11	3075	3	4643	7
13	16	3	2643	12	3993	14
14	15	7	2910	9	4395	10
15	13	12	2017	14	4123	12

\* X 10,000

---LTF---

Informante	Rango		dc(i) * LL2-0.75		dc(i) * LS	
1	11	13	1854	12	3240	12
2	18	7	2300	5	4388	4
3	20	3	2454	2	4744	1
4	11	15	1670	13	2586	15
5	16	10	2094	9	3857	9
6	19	6	2109	7	3984	7
7	19	5	2411	3	4647	3
8	22	1	2463	1	4727	2
9	20	2	2109	8	3869	8
10	15	11	1241	15	2609	14
11	11	14	1876	11	3360	11
12	16	9	2250	6	4027	6
13	11	12	2020	10	3513	10
14	17	8	1416	14	2881	13
15	20	4	2322	4	4265	5

\* X 10,000

---LCF---

Informante	Rango		dc(i) * LL2-0.75		dc(i) * LS	
1	12	15	3435	12	5406	11
2	16	4	3627	7	5659	8
3	16	3	3369	13	5394	12
4	16	2	3686	3	5851	3
5	13	12	3320	14	5003	14
6	15	7	3500	9	5270	13
7	14	11	3442	11	5432	9
8	15	6	3834	1	6036	1
9	15	9	3649	6	5742	6
10	17	1	3700	2	5892	2
11	14	10	3675	4	5754	5
12	13	13	3189	15	4918	15
13	16	5	3671	5	5805	4
14	13	14	3457	10	5415	10
15	15	8	3610	8	5732	7

\* X 10,000

---LHF---

Informante	Rango		dc(i) * LL2-0.75		dc(i) * LS	
1	21	8	2360	6	5516	2
2	23	3	2664	1	5995	1
3	23	2	2305	9	5248	7
4	19	11	2430	4	5300	6
5	17	12	2390	5	5158	8
6	17	13	2309	8	4926	10
7	21	7	2433	3	5461	3
8	21	5	2122	14	4573	13
9	15	15	1473	15	2967	15
10	22	4	2448	2	5129	9
11	20	9	2253	10	5324	4
12	16	14	2342	7	4758	12
13	19	10	2200	12	4233	14
14	24	1	2163	13	4850	11
15	21	6	2216	11	5322	5

-----  
\* X 10,000

Orden entre informantes por centro

Informante	A	N	L	J	R	T	C	H
1	14	5	13	8	15	12	12	6
2	3	14	9	14	10	5	7	1
3	10	11	2	10	7	2	13	9
4	15	15	1	11	13	13	3	4
5	12	3	4	15	5	9	14	5
6	9	7	5	9	2	7	9	8
7	6	1	3	7	11	3	11	3
8	5	2	11	4	6	1	1	14
9	8	12	14	6	1	8	6	15
10	1	13	10	5	8	15	2	2
11	11	4	7	2	4	11	4	10
12	2	8	8	12	3	6	15	7
13	7	6	15	1	12	10	5	12
14	13	10	6	3	9	14	10	13
15	4	9	12	13	14	4	8	11

APENDICE G

DISPONIBILIDAD DE INDIVIDUOS

APLICACION DE DOS FORMULAS

Informante	d1(i) *		d2(i) *	
	LL2-0.75		LL2-0.75	
1	16412	15	1991	15
2	18835	8	2266	8
3	19626	3	2383	3
4	18558	10	2196	10
5	17509	14	2068	13
6	19424	5	2340	6
7	20418	1	2511	1
8	19039	7	2271	7
9	18264	11	2184	11
10	19807	2	2420	2
11	18650	9	2215	9
12	19252	6	2364	5
13	19500	4	2374	4
14	17546	13	2065	14
15	17759	12	2169	12

-----  
\* X 10,000

## BIBLIOGRAFIA

- Alarcos LLorach, E., Gramática Estructural, Madrid, Gredos, 1969.
- Alvarez Nazario, M., "El Vocabulario Campesino Puertorriqueño que Recoge la Literatura Costumbrista del XIX", Atenea, 1981, 2, 15-27.
- Arnold, H. H., "A List of Graded Vocabularies and Method of Grading", Modern Language Journal, 1932, 16, 644-655.
- Avila, R., "Léxico Infantil de México: Palabras tipos y vocablos", Actas de II Congreso Internacional sobre el Español de América, ed. , José G. Moreno de Alba, Universidad Autónoma de México, Facultad de Filosofía y Letras, 1986, 510-517.
- \_\_\_\_\_, G. Ruiz de Bravo Ahuja y B. Galazar Cuarón, Léxico Infantil de México: Recopilación Preliminar, México, 1972.
- Azorín, F. P., Curso de Muestreo y Aplicaciones, Madrid, Instituto Nacional de Estadística, 1962.
- Azurmendi Ayerbe, M. J., Elaboración de un Modelo para la Descripción Sociolingüística del Bilingüismo y su Aplicación Parcial en la Comarca de San Sebastián, Guipúzcoa, Caja de Ahorros Provincial de Guipúzcoa, I, II y III, 1983.
- Bloomfield, L., Language, New York, Holt and Co., 1946.
- Bocaz, A., "Análisis de las Denominaciones de Vehículos en Santiago de Chile", Revista de Lingüística Teórica y Aplicada, 1972, 10, 89-105.
- Bortolino, V. et al., Lessico di Frequenza della Lingua Italiana Contemporanea, Italia, IBM, 1972.
- Bright, W., "Introduction: The Dimensions of Sociolinguistics" Sociolinguistics, Proceedings of the UCLA Sociolinguistics Conference, La Haya, Mouton, 1966.
- Buchanan, M., A Graded Spanish Work Book, Anual C. E. Modern Languages, Toronto, Toronto University Press, 1941, III.
- Canalleda, M. J., "Sobre Lenguaje Infantil", Filología, 1968-1969, 1970, 13, 39-47.
- Canizal Arévalo, A. V., Disponibilidad Léxica en Escolares de Primaria Terminada. Análisis de 6 Centros de Interés, tesina inédita, México, Universidad Nacional Autónoma de México, 1987.



Cartwright, C. W., "A Study of the Vocabularies of Eleven Spanish Grammar and Fifteen Spanish Reading Texts", Modern Language Journal, 1925, 10, 1-14.

Céspedes, T. R., Investigación Acerca de las Palabras Usadas en Castellano, Panamá, Star and Herald, 1928-29.

Chase, C. I., Elementary Statistical Procedures, New York, Mc Graw-Hill, 1984.

Deming, W. E., Some Theory of Sampling, New York, John Wiley and Sons, 1950.

Dimitrijévich, N., Lexical Availability, Heidelberg, Julius Gross Verlag, 1969.

Dubois, J. "Distribution, Ensemble et Marque dans le Lexique", Cahiers de Lexicologie, 1966, 2, 62-66.

Echeverría, M. S., "Programa Computacional Interactivos en Lingüística", Revista de Lingüística Teórica y Aplicada, 1985, 23, 139-142.

\_\_\_\_\_, Disponibilidad Léxica en Educación Media, Concepción, Chile, Universidad de Concepción, 1987.

\_\_\_\_\_ y M. Olivia Herrera, "Incidencia de Variables Socioculturales en la Comprensión de Vocabulario y Sintaxis del Español", ponencia presentada al VI Congreso de Alfal, Phoenix, Arizona, Sept. 1981.

Fishburn, P. C. y R. L. Keeney, "Seven Independence Concepts and Continuous Multiattribute Utility Functions", Journal of Mathematical Psychology, 1974, 11, 294-327.

Galisson, R., Lexicologie et Enseignement des Langues, París, Hachette, 1979.

García Hoz, V., "Evolución Cuantitativa del Vocabulario en Escolares de Nueve a Dieciocho Años", Revista Española de Filología, 1946, 16, 403-433.

\_\_\_\_\_, Vocabulario Usual, Vocabulario Común y Vocabulario Fundamental, Madrid, CSIC, 1953.

Garvin, R. L., ed., Natural Language and the Computer, New York, Mc Graw Hill, 1963.

Garvin, P., "La Lingüística y la Enseñanza de la Lengua Materna", Revista de Lingüística Teórica y Aplicada, 1972, 10, 27-48.

- Gili Gaya, S., Estudios de Lenguaje Infantil, Barcelona, Bibliograph, S.A., 1974.
- \_\_\_\_\_, Nuestra Lengua Materna, San Juan, Puerto Rico, 1965.
- \_\_\_\_\_, Curso Superior de Sintaxis Española, Barcelona, 1961.
- González, A., "Léxico Básico de la Lengua Escrita en República Dominicana y Conocimiento que del Mismo Tienen los Estudiantes al Finalizar la Educación Media", Actas del II Congreso Internacional sobre el Español de América, ed., José G. Moreno de Alba, Universidad Nacional Autónoma de México, Facultad de Filosofía y Letras, 1986, 523-531.
- Gougenheim, G., "La Statistique de Vocabulaire et son Application dans L'Enseignement des Langues", Revue de L'Enseignement Supérieur, 1967, 3, 137-144.
- Gross, M., Modelos Matemáticos en Lingüística, Madrid, Gredos, 1976.
- Guilbert, L., L'Utilisation de la Statistique en Lexicologie Appliquée, Etudes de Linguistique Appliquée, Université de Besançon, Faculté des Lettres et Sciences Humaines, 2, 12-23.
- Guiraud, P., Caractères Statistiques du Vocabulaire, París, Presses Universitaires de France, 1954.
- \_\_\_\_\_, "Statistique Linguistique et Histoire du Vocabulaire", Cahiers de Lexicologie, 1965, 2, 37-40.
- \_\_\_\_\_, et al., L'Elaboration du Français Fondamental (1er. degré), París, Didier, 1964.
- \_\_\_\_\_, "Les Enseignements de la Statistique de Vocabulaire", Etudes de Ling. Appliquée, 1963, 2, 5-11.
- Halliday, M., Language as Social Semiotic. The Social Interpretation of Language and Meaning, London, Edward Arnold, 1978.
- Harris, Z., Structural Linguistics, Chicago, University of Chicago, 1960.
- Henmon, V. A. C., A French Work Book Based on the Count of 400,000 Running Words, Madison, Bureau of Educational Research, University of Wisconsin, 1924.
- Hernández Aquino, L., "Lexicografía Indígena de Puerto Rico", Boletín de la Academia Puertorriqueña de la Lengua Española, 1980, VIII<sub>1</sub>, 25-50.

Hillier, F. y G. Lieberman, Introducción a la Investigación de Operaciones, México, Mc Graw Hill, 1982.

Huyke, I., "Índices de Densidad Léxica: Anglicismos en la Zona Metropolitana de San Juan", Corrientes Actuales en la Dialectología del Caribe Hispánico, Río Piedras, Editorial Universitaria, 1978.

Jakobson, R., "Linguistics and Communication Theory", Structure of Language and its Mathematical Aspects, Providence, American Mathematical Society, 1961, 245-252.

Jameison, E. J., "A Standardized Vocabulary of Elementary Spanish", Modern Language Journal, 1924, 8, 325-333.

Juilland, A. y V. Traversa, Frequency Dictionary of Italian Words, The Hague, Mouton, 1973.

Juilland, A. y A. Roceric, The Linguistic Concept of Word, Janua Linguarum Series Minor, 130, La Hague, Mouton, 1972.

Juilland, A. et al., Frequency Dictionary of French Words, The Hague, Mouton, 1971.

\_\_\_\_\_, Frequency Dictionary of Rumanian Words, The Hague, Mouton, 1965.

Juilland, A. y E. Chang-Rodríguez, Frequency Dictionary of Spanish Words, The Hague, Mouton, 1964.

Justo Hernández, H. G., Disponibilidad Léxica en Colores, tesina inédita, México, Universidad Nacional Autónoma de México, 1986.

Kaeding, J. W., Haufigkeitswörterbuch der Deutschen Sprache, Steglitz bei Berlin, der Herausgeber, 1897.

Keeney, R. y H. Raiffa, Decisions with Multiple Objective, New York, John Wiley and Sons, 1975.

Keniston, H., "Common Words in Spanish", Hispania, 1920, 3, 85-96.

\_\_\_\_\_, Basic List of Spanish Words and Idioms, Chicago, The University of Chicago Press, 1928.

\_\_\_\_\_, Spanish Idiom List, Selected on the Range and Frequency of Occurrence, New York, The Macmillan Co., 1929

\_\_\_\_\_, A Standard List of Spanish Words and Idioms, New York, Heath and Co., 1941.

Kish, L., Muestreo de Encuestas, México, Trillas, 1972.

Labov, W., Sociolinguistic Patterns, Filadelfia, University of Pennsylvania Press, 1972b.

\_\_\_\_\_, "Some Principles of Linguistic Methodology", Language and Society, 1972, I, 12-19.

Lakoff, R., Language and Woman's Place, New York, Harper Colophon Books, 1975.

Lamb, S. M., "The Digital Computer as an Aid in Linguistic Research", Language, 1962, 38, 385-389.

Lamíquiz, V., "Sociolingüística en un Habla Urbana: Sevilla", RSEL, 1976, Fasc.2.

Lara, L. F. et al., Investigaciones Lingüísticas en Lexicografía, México, El Colegio de México, 1979.

Lope Blanch, J. M., El Español de América, Madrid, Alcalá, 1968.

López Chávez, J. y C. Strassburger Frías, Otro Cálculo del Índice de Disponibilidad Léxica, México, Universidad Autónoma de México, 1987.

López Morales, H. y R. Lorán, "Nouveau Calcul de L'Indice de Disponibilité", (en prensa), 1983.

López Morales, H., Introducción a la Lingüística Actual, Madrid, Playor, 1983.

\_\_\_\_\_, Enseñanza de la Lengua: Lingüística para Maestros de Español, Madrid, Playor, 1984.

\_\_\_\_\_, "Disponibilidad Léxica y Estratificación Socioeconómica", Dialectología y Sociolingüística: Temas Puertorriqueños, Madrid, Playor, 1984, 173-181.

\_\_\_\_\_, "Frecuencia Léxica, Disponibilidad y Programación Curricular", Boletín de la Academia Puertorriqueña de la Lengua Española, 1978, VI<sub>1</sub>, 73-86.

\_\_\_\_\_, "Frecuencia y Disponibilidad Léxica en Escolares de Primer Grado", ponencia presentada al IV Congreso de ALFAL, enero de 1975.

Lorán, R., La Tasa de Sustitución y el Índice de Disponibilidad Léxica, Comunicación Técnica, Depto. de Matemáticas, Recinto Universitario de Mayagüez, Universidad de Puerto Rico, 1987.

- \_\_\_\_\_, Un Índice de Disponibilidad Léxica, Comunicación Técnica, Depto. de Matemáticas, Recinto de Mayagüez, Universidad de Puerto Rico, 1983.
- Lorenzo, E., El Español de Hoy, Lengua en Ebullición, Madrid, Gredos, 1971.
- Lotus 123 Manual, Lotus Development Corporation, 1986.
- Lyons, J., Semántica, Barcelona, Teide, 1980.
- Mackey, W. F., Le Vocabulaire Disponible du Français, I, París, Didier, 1971.
- \_\_\_\_\_, Le Sondage dans les Enquêtes de Disponibilité, Québec, CIRB, 1971.
- Mena Osorio, M., Disponibilidad Léxica Infantil en Tres Niveles de Enseñanza Básica, tesina inédita, Concepción, Chile, Universidad de Concepción, 1986.
- Mendenhall, W., Introducción a la Probabilidad y la Estadística, California, Wadsworth International, 1982.
- Michéa, R., "Introduction Pratique a une Statistique du Langage", Les Langues Modernes, 1969, 173-186.
- \_\_\_\_\_, "La Notion de Situation en Linguistique Appliquée", Etudes de Linguistique Appliquée, 1967, 9-15.
- \_\_\_\_\_, "Mots Fréquents et Mots Disponibles. Un Aspect Nouveau de la Statistique du Langage", Les Langues Modernes, 1953, 47, 338-344.
- \_\_\_\_\_, "Vocabulaire et Culture", Les Langues Modernes, mayo-junio, 1950.
- Milton, J. S. et al., Introduction to Statistics, New York, D. C. Heath and Company, 1986.
- Morales, A., Léxico Básico del Español de Puerto Rico, San Juan, Academia Puertorriqueña de la Lengua, 1986.
- \_\_\_\_\_, Teorías y Métodos Aplicables a la Estadística Lingüística del Español Americano, tesis inédita, Río Piedras, Universidad de Puerto Rico, 1975.
- Moureau, M., Bulletin d'Information du Laboratoire d'Analyse Lexicologique, 6, Centre d'Etude du Vocabulaire Français de la Faculté des Lettres et Sciences Humaines de Besançon.
- Muller, Ch., Estadística Lingüística, Madrid, Gredos, 1967.

- \_\_\_\_\_, "Frequence, Dispersion et Usage", Cahiers de Lexicologie, 1965, 7, 33-42.
- \_\_\_\_\_, "Sur la Mesure de la Richese Lexicale Théorie et Expériences", Etudes de Linguistique Appliquée, 1971, 20-46.
- Njock, P. E., L'Universe Familier de l'Enfant Africain, Québec, CIRB, 1979.
- Ot, L., An Introduction to Statistical Methods and Data Analysis, Massachusetts, Duxbury Press, 1977.
- Pandolfi, A. M. y M. O. Herrera, "Análisis Léxico-Semántico en el Niño Menor de Tres Años", Revista de Lingüística Teórica y Aplicada, 1985, 23, 121-131.
- Parada, E., "El Léxico Figurado de las Minas de Carbón-Lota", Revista de Lingüística Teórica y Aplicada, 1974, 12, 15-36.
- Peronard, M., "Variación Diastrática y Lenguaje Infantil", Boletín de la Academia Puertorriqueña de la Lengua Española, 1978, VI<sub>1</sub>, 107-126.
- Personal Computer Software NCR-DOS, Washington, Microsoft Corporation, 1986.
- Phillips, G., "Lingüística Computacional", en Revista de Lingüística Teórica y Aplicada, 1980. 18, 155-176.
- Ramírez Gasca, R., "Uso de Anglicismos en el Habla Estudiantil", Actas del II Congreso Internacional sobre el Español de América, ed., José G. Moreno de Alba, Universidad Nacional Autónoma de México, Facultad de Filosofía y Letras, 1986, 589-592.
- Rodríguez Bou, I., Recuento de Vocabulario de Pre-Escolares, Río Piedras, Consejo Superior de Enseñanza, 1966, I y II.
- \_\_\_\_\_, Recuento de Vocabulario de Revistas, Río Piedras, Consejo Superior de Enseñanza, 1958.
- \_\_\_\_\_, "La Lengua Hablada en la Escuela Elemental", en Publicaciones Pedagógicas, Río Piedras, Consejo Superior de Enseñanza, 1952.
- \_\_\_\_\_, Recuento de Vocabulario Español, Río Piedras, Consejo Superior de Enseñanza, Waverly, 1952, I.
- Román, B., Disponibilidad Léxica en Escolares de Dorado, tesina inédita, Río Piedras, Universidad de Puerto Rico, 1985.

- Russell, H. J., The Most Common Spanish Words and Idioms, New York, The Macmillan Co., 1937.
- Urrutibéheity, H. N., "The Statistical Properties of the Spanish Lexicon", Cahiers de Lexicologie, 1971, 19, 79-95.
- Uthley, J. H., "A Mexican Word List", Hispania, 1940, 23, 857-861.
- Vaquero de Ramírez, M. T., "Enseñar Español, pero ¿Qué Español?", Boletín de la Academia Puertorriqueña de la Lengua Española, 1978, VI<sub>1</sub>, 127-146.
- Villamil de Forastieri, B., "El Desarrollo de la Lengua Materna y su Relación con el Aprendizaje de la Lectura", Boletín de la Academia Puertorriqueña de la Lengua Española, 1978, VI<sub>1</sub>, 147-161.
- Wade, G., "A Note on Spanish Vocabulary", Modern Language Journal, 1938, 22, 433-436.
- Word Perfect Manual, Version 4.1, Utah, Satellite Software International, 1986.
- Zierer, E., "Métodos Matemáticos en la Lingüística" Simposio de San Juan, Puerto Rico, 1971, PILEI, San Juan, Departamento de Instrucción Pública, 1974, 133-140.
- Zipf, G. K., "Statistical Method and Dynamic Philology", Language, 1937, 13, 60-69.

## TABLA DE CONTENIDO

	<u>Página</u>
INTRODUCCION . . . . .	1
CAPITULO I : LOS ESTUDIOS CUANTITATIVOS . . . . .	5
1. Generalidades . . . . .	5
2. Antecedentes en el tema: la léxico- estadística . . . . .	6
3. Más antecedentes: la disponibilidad léxica . . . . .	15
4. Nuestra investigación . . . . .	20
5. Organización del presente trabajo . . . . .	24
6. Recuento bibliográfico . . . . .	25
CAPITULO II : LOS INDICES DE DISPONIBILIDAD . . . . .	33
1. Introducción . . . . .	33
2. El índice LL1 . . . . .	37
3. El índice LL2 . . . . .	41
4. El índice LS . . . . .	45
5. Comparación entre los diferentes índices . . . . .	47
6. Procesamiento de los datos . . . . .	50
CAPITULO III : NUEVOS INDICES DE DISPONIBILIDAD . . . . .	56
1. Introducción . . . . .	56
2. Disponibilidad de centros para un grupo . . . . .	59
3. Disponibilidad léxica de un informante . . . . .	62
4. Disponibilidad léxica de un grupo . . . . .	67
CAPITULO IV : ASPECTOS SOCIOLINGUISTICOS . . . . .	72
1. Introducción . . . . .	72
2. Lengua hablada; lengua escrita . . . . .	74
3. La encuesta . . . . .	75
4. Listas abiertas y listas cerradas . . . . .	78
5. Comparaciones entre grupos . . . . .	79
CAPITULO V : CENTROS DE INTERES . . . . .	83
1. Introducción . . . . .	83
2. Antecedentes . . . . .	83



3. Centros de interés en nuestra investigación . . . . .	90
4. Segmentación de la unidad léxica . . . . .	95
5. Índice de cohesión . . . . .	97
CONCLUSIONES . . . . .	100
Resultados de la comparación de fórmulas . . . . .	100
Sobre nuevos índices . . . . .	102
Sobre la metodología del procesamiento de datos . . . . .	103
Investigaciones futuras . . . . .	104
APENDICE A . . . . .	106
APENDICE B . . . . .	120
APENDICE C . . . . .	126
APENDICE D . . . . .	128
APENDICE E . . . . .	135
APENDICE F . . . . .	146
APENDICE G . . . . .	156
BIBLIOGRAFIA . . . . .	158
TABLA DE CONTENIDO	
TABLA DE CUADROS	

TABLA DE CUADROS

		<u>Página</u>
Cuadro 1	Diversos tipos de índices. . . . .	57
Cuadro 2	Posibles caminos para el cálculo de los diferentes índices. . . . .	58
Cuadro 3	Disponibilidad de grupo para diferentes centros según diferentes fórmulas . . . . .	62
Cuadro 4	Cómputo de la fórmula $D3_C(g)$ para las escuelas en cada centro. . . . .	80
Cuadro 5	Cómputo de la fórmula $D3_C(g)$ para las escuelas (grupos masculinos) en cada centro de interés. . . . .	80
Cuadro 6	Cómputo de la fórmula $D3_C(g)$ para las escuelas (grupos femeninos) en cada centro de interés. . . . .	81
Cuadro 7	Cómputo de la fórmula $D6(g)$ para las diferentes escuelas. . . . .	82
Cuadro 8	Centros de interés empleados por diferentes investigadores. . . . .	92
Cuadro 9	Índice de cohesión para todos los centros del grupo femenino de la escuela "La Merced". . . . .	98
Cuadro 10	Índice de cohesión para el centro "alimentos" de los grupos femeninos. . .	99