Web personal: http://angarmegia.tresuvesdobles.com - http://angarmegia.95mb.com
Correo electrónico: angarmegia@telefonica.net

La ponderación de la posición y los coeficientes de DISPONIBILIDAD LÉXICA

Conveniencia de ensayar otros procedimientos de cálculo

El modelo probabilístico desarrollado por Antonio García Megía

Primera Parte

Para profundizar en el tema de este: García Megía, Antonio.. *La disponibilidad léxica en la ciudad de Almería entre los nueve y los doce años*. Servicio de Publicaciones. Universidad de Almería. Almería, 2004

Un somero ejercicio analítico sobre los algoritmos de cálculo que se aplican en la actualidad para estimar los *coeficientes de disponibilidad*, basados en las formulaciones de López Morales y Lorán modificados posteriormente por López Chávez y Strarburger Frías, deja claro que establecen sus conclusiones a partir de la combinación de dos factores: frecuencia y posición, referidos ambos a los vocablos contenidos en las encuestas de recuerdo libre que producen los individuos que proporcionan la muestra que maneja el investigador y que constituyen la base del entramado metodológico de la *disponibilidad*. La frecuencia de aparición de cada vocablo en el conjunto de la población estudiada queda matizada por el lugar o posición en que es recordado por todos y cada uno de los sujetos que lo evocan.

En este contexto, el tratamiento y consideración del elemento *posición* resulta esencial para aceptar y validar como buenos los guarismos obtenidos.

Los principios que caracterizan la manera actual de operar para establecer los *índices de disponibilidad* se apoyan en las siguientes premisas establecidas por los maestros antes citados:

1. Se declaran unas condiciones previas, *axiomas*, que condicionan y determinan las formulas de aplicación.
2. Se asignan *pesos* o coeficientes compensadores que ponderan y otorgan valor a

Web personal: http://angarmegia.tresuvesdobles.com - http://angarmegia.95mb.com
Correo electrónico: angarmegia@telefonica.net

cada una de las *posiciones* que ocupan los vocablos en las listas, convirtiendo así los datos ordinales en medidas de intervalo.

3. Se aceptan y cuantifican valores constantes que se incorporan al cálculo para conseguir distribuciones de *pesos* para las *posiciones* que correspondan a una función exponencial definida en un rango en abscisas entre *1* y *0.1002*.
4. Se trasladan, trabajan y estudian, el número de las menciones en cada *posición* sobre una matriz. Los cálculos se realizan a partir de sus *frecuencias marginales.*

El objeto de este trabajo es llamar la atención sobre determinados aspectos del método y animar al colectivo investigador a encontrar caminos alternativos al modelo aceptado para corregir ciertas disfunciones. En este sentido permítase aducir algunas razones.

La declaración axiomática que enmarca cualquier investigación debe emanar siempre de la naturaleza de los datos que se consideran y no de los juicios apriorísticos de quienes han de analizarlos. Sirva de ejemplo la condición impuesta por López Chávez y Strassburger Frías [LÓPEZ CHÁVEZ y STRASSBURGER. 1987:20-21] que exige del *coeficiente de dispersión* el requisito:

[*Posición* i] + [*Posición* (i+2)] > 2 [*Posición* (i+1)]

donde *i* es *posición*

Si se advierte que la investigación en *disponibilidad léxica* trabaja a partir de listas con el valor de *escalas ordinales*, y que el *orden* indica *posición relativa*, pero *nunca mide la magnitud que separa una posición de otra*, la aplicación indiscriminada de ese principio puede desvirtuar la realidad de los hechos.

Sean los puntajes *a = 1*; *b = 0.9* y *c = 0.5,* con ellos se confecciona la tabla que sigue:

Posición i	**Posición (i+1)**	**Posición (i+2)**
a	b	c
1	0.9	0.5

En esta situación hipotética resulta evidente que otorgar valores a los puestos según los criterios de la premisa comentada altera la verdad de los datos, toda vez que

Web personal: http://angarmegia.tresuvesdobles.com - http://angarmegia.95mb.com
Correo electrónico: angarmegia@telefonica.net

(1 + 0.5) no es mayor que (2 * 0.9)

Se puede argumentar que se opera con elementos nacidos de actos volitivos entre los que no es posible establecer con total precisión correspondencias numéricas. Y es así, pero cuando se pretende elaborar una teoría y deducir conclusiones consecuentes a las diferencias detectadas acudiendo a recursos matemáticos y estadísticos, éstos han de respetarse íntegramente en sus reglas y servidumbres para que las aseveraciones consiguientes sean aceptadas, sin sombra de duda, como científicas. Advierte Jesús Amón al estudioso de las ciencias de la conducta que reclama el auxilio de modelos matemáticos:

> "Lo verdaderamente importante es que sepa acercarse con mentalidad matemática a los problemas que se le presenten. Es decir, que se esfuerce por asimilar el proceso lógico subyacente al modelo matemático de que se trate, que conozca las condiciones que lo hacen posible y, consiguientemente, las condiciones que éste exige de la realidad concreta para que sea posible su aplicación a la misma". [AMON, J. 1990:23].

Es legítimo el uso de instrumentos específicos de las ciencias exactas en actuaciones desarrolladas en el ámbito de las ciencias de la conducta siempre que las mediciones cumplan determinadas condiciones.

La definición más aceptada de medida, "la asignación de números a objetos y sucesos de acuerdo con reglas lógicamente aceptables", se atribuye a Campbell [KOHAN, N, y CARRÓ, J. 1968]. Existe la posibilidad de describir válidamente cualquier realidad mediante números siempre que haya un grado suficiente de *isomorfismo*, entendido como semejanza de propiedades o forma. Las propiedades que pueden poseer conjuntamente fenómenos y números son las de *identidad*, *orden* y *aditividad*.

> "Todo número es único y ningún otro es exactamente lo mismo. Un número tiene pues *identidad*. Así que todo objeto o suceso al que se aplique un número debe tener también *identidad*. En el sistema numérico, los números tienen la propiedad del *orden* o *rango*, pues de dos números uno es mayor que el otro. Los objetos a los que se aplican deben ser ordenables según un continuo dado, si el *orden* de los números asignados a ellos ha de suministrar una descripción de *orden*. Los números también tienen la propiedad de la *aditividad*, lo cual quiere decir que la suma de un número determinado con un número determinado debe dar invariablemente un número único". [GUILFOR y FRUCHTER. 1984:18-19].

No es imperativo que los fenómenos a los que se aplican números las posean todas ellas para que resulten mensurables, pero la utilidad de los números aplicados a la medición depende de cuántas de esas propiedades les son de referencia. Son los atributos inherentes a las cifras que representan los hechos los que definen la escala de medición pertinente y, en consecuencia, el nivel de operaciones y abstracciones permitido. Lo que se demanda de quien utiliza números para alcanzar conclusiones es la selección exquisita de los procedimientos para adecuarlos a la tipología de sus variables [GUILFOR y FRUCHTER. 1984:11-23].

Es esencial, por ejemplo, adecuar los cálculos matemáticos y estadísticos a la clase de datos con los que se opera. Siegel [SIEGEL. 1983:51], define las relaciones que se admiten dentro de cada categoría y muestra en esquema las operaciones permitidas para cada rango de datos. De acuerdo con ella no es procedente, por ejemplo, calcular una *media geométrica* a partir de datos de intervalo. De igual manera determinar una *media natural* cuando se opera con medidas de proporción, que admiten la *media geométrica*, puede significar una perdida de información adicional valiosa para el proyecto en que se incardina la estimación.

Escala	Relaciones definidas	Estadísticos apropiados	Pruebas apropiadas
Nominal	Equivalencia	Moda Frecuencia Coeficiente de contingencia	Pruebas estadísticas no paramétricas
Ordinal	Equivalencia De mayor a menor	Mediana Percentiles Spearman r_s Kendall r	
Intervalo	Equivalencia De mayor a menor Proporción conocida de un intervalo a cualquier otro	Media Desviación estándar Correlación del momento-producto de Pearson. Correlación del múltiple momento producto	Pruebas estadísticas paramétricas y no paramétricas.
Proporción	Equivalencia De mayor a menor Proporción conocida de un intervalo a cualquier otro Proporción conocida de un valor de la escala a cualquier otro	Media geométrica Coeficiente de variación	

En el campo de la *disponibilidad léxica*[1] se atiende a variables de tipo *discreto,* no existen recuentos fraccionarios, *nominal,* los datos recogidos no son cuantitativos, y de

[1] Ver los capítulos dedicados a lengua y estadística en García Megía, A. 2004 y la misma temática y autor en http://angarmegia.treuvesdobles.com o http://angarmegia.95mb.com

Web personal: http://angarmegia.tresuvesdobles.com - http://angarmegia.95mb.com
Correo electrónico: angarmegia@telefonica.net

categoría infinita, pueden aparecer más de veinte respuestas diferentes. Es lícito, en consecuencia, operar con aquellos cálculos que les son compatibles, pero no con otros, tal vez más potentes, pero sólo aptos para *escalas de intervalo* o superiores. Les son de aplicación ejercicios de clasificación y recuento. No permiten, en cambio, la adición ni la sustracción [FOX, D. J. 1981:173-202].

Otro aspecto a considerar es la posible distorsión de la objetividad de los hechos que, en virtud de las constantes utilizadas, puede derivarse de la asignación de *pesos* a *rangos* para incrementar la potencia de las escalas ordinales.

No quiere decir esto que se no sea lícito el procedimiento, lo que se subraya es que el método debe articularse *sólo* a partir de la información que suministran los propios datos. Son ellos, *los datos* que reflejan la realidad y conforman su retrato matemático en virtud del *isomorfismo* existente entre el fenómeno observado y el número que lo representa, los únicos legitimados para establecer los límites y pautas de su metamorfosis.

Por otra parte, centrar los cálculos en las *frecuencias* marginales de las matrices que resumen los recuentos de vocablos y posiciones, y atribuir valores a las *tasas de sustitución* y *coeficientes de dispersión*, ajustados entre *1* y *0,1002* para facilitar la comparación de muestras distintas, puede olvidar un hecho de relevancia: la *disponibilidad* de la palabra que recuerda en primer lugar un individuo capaz de enunciar un listado de treinta vocablos debe ser mayor, en función de la amplitud de su repertorio, que la misma *posición* en otro informante que sólo relaciona diez términos.

La significación de cualquier fenómeno queda establecida en razón inversa a la probabilidad de su aparición fortuita. Sobre este principio se contrasta, acepta o rechaza, toda tesis científica[2]. De igual manera se podría afirmar que la fuerza que asocia una respuesta a un determinado estímulo es tanto más consistente cuanto menos esperable es en un universo regido por las leyes del azar. Análogamente, si la *disponibilidad léxica* pretende medir la intensidad de la atracción que mueve al hablante a actualizar un término en un contexto concreto, se entiende que el poder de seducción de una situación para una determinada expresión, aquel que provoca su elección entre varias, será tanto más profundo cuanto mayor sea el vocabulario a disposición del sujeto.

La propuesta metodológica alternativa que aquí se defiende habrá de estructurarse

[2] El contraste de hipótesis analiza la probabilidad de ocurrencia de la hipótesis de nulidad (H_0) y de la hipótesis alternativa (H_1), es una aportación esencial de la Estadística a la investigación.

Web personal: http://angarmegia.tresuvesdobles.com - http://angarmegia.95mb.com
Correo electrónico: angarmegia@telefonica.net

teniendo en cuenta los argumentos expuestos, lo cual implica:

1. La valoración del alcance significativo de las posiciones de cada lista en función de su *rango* o longitud. Ello relega a un segundo término el papel de la *frecuencia* marginal y centra el estudio de la *disponibilidad* en el análisis de las respuestas de los informantes considerados individuo a individuo. La necesaria interpretación de los resultados dentro del conjunto del colectivo encuestado corresponde a un momento posterior.
2. Los cálculos deben estructurarse a partir de teoremas, modelos y métodos matemáticos aceptados universalmente y aplicables al tipo de variables y objetivos del proyecto.

Un modelo matemático construido a partir de la teoría y cálculo de probabilidades[3] puede ajustarse con ventaja a tales premisas. La teoría de las probabilidades se aplica universalmente a la investigación, bien como base para la elaboración de un método o modelo inferencial, bien como instrumento de validación o rechazo de hipótesis científicas. Ya se ha repetido que su utilidad nace de la capacidad que demuestra para evaluar la significación de una situación comparando los resultados obtenidos y los esperables por azar. En este caso, reduce el acto de establecer la probabilidad de la mención asociada a la *posición* a un cálculo elemental.

Dice Guilford, aunque la definición se remonta a Laplace:

> "La probabilidad de que un suceso se presente de la manera que nos interesa (suceso favorable) es el cociente del número de maneras en que ese suceso favorable puede ocurrir por [entre] el número total de maneras como el suceso puede ocurrir". [GUILFORD, J. y FRUCHTER, B. 1984:89].

Otro valor destacable radica en su posibilidad de uso con diferentes clases de datos y *escalas de medida*.

La tabla que sigue recoge el supuesto caso de un informante que confecciona una relación de seis palabras.

[3] Existen innumerables fuentes para profundizar en el mundo de la probabilidad. Sólo a modo de ejemplo se sugieren: LAPLACE [1812; 1825], CARRANZA [1961], MORA [1989], RIVADULLA [1991]...

Web personal: http://angarmegia.tresuvesdobles.com - http://angarmegia.95mb.com
Correo electrónico: angarmegia@telefonica.net

Ejemplo de cálculo de la probabilidad de elección de cada palabra en la *posición* indicada

Probabilidad: $\Delta = \mathbf{N}_{\text{favorables}} / \mathbf{N}_{\text{posibles}}$						
	Posición 1	Posición 2	Posición 3	Posición 4	Posición 5	Posición 6
Mención	a	b	c	d	e	f
$N_{favorable}/N_{posible}$ [4]	1/6	1/5	1/4	1/3	1/2	1/1
Probabilidad	0.16667	0.2	0.25	0.33333	0.5	1

Antes de seguir adelante es conveniente recordar y fijar algunos conceptos básicos [GUILFORD, J. y FRUCHTER, B. 1984:87-102]:

- De la definición de probabilidad asumida se deduce que la cuantificación de la posibilidad aleatoria de ocurrencia de un suceso se mueve entre los valores 0, del suceso imposible, y 1, del suceso seguro.
- La suma de las probabilidades de todos los sucesos posibles ha de ser igual a *1*.
- Se entiende por *probabilidad complementaria* de un suceso la que corresponde a la suma de probabilidades de todos los demás. Se calcula restando de *1* la probabilidad del suceso en cuestión.

Aunque en este lugar no se pretende desarrollar detalladamente el modelo metodológico para el cálculo de la *disponibilidad léxica* incardinado en este universo matemático, que es objeto de otro trabajo, sirvan algunas consideraciones para apuntar las líneas de investigación que se han seguido y que lo sustentan.

Si la *disponibilidad* aspira a medir la intensidad de la relación que asocia tema y palabra, se puede convenir que tal vínculo ha de vencer la resistencia que el azar impone para que se actualice una determinada elección entre varias posibles. Es decir, *la fuerza desplegada por la voluntad del hablante, por el inconsciente que evoca el término más familiar o la empatía que encadena vocablo y estímulo, debe, cuando menos, igualar la potencia de sentido contrario que desarrollan las leyes reguladoras de la casualidad, única responsable, caso de no incidir ninguna circunstancia externa, de que se lleve a efecto un determinado acontecimiento.*

La tabla anterior calcula la probabilidad para que "*a* se mencione en primer lugar" en *0.16***.** Consecuentemente, el suceso contrario, "*a* no se menciona en primer lugar", se estima

[4] Cuando el hablante menciona la primera palabra dispone de seis posibilidades, pero cuando elige la segunda selecciona entre cinco, ya que no puede repetir la primera.

Web personal: http://angarmegia.tresuvesdobles.com - http://angarmegia.95mb.com
Correo electrónico: angarmegia@telefonica.net

en *0.83*[5]. Tal valor podría servir de punto de partida para deducir el indicador de la *disponibilidad* que se busca. Se debe notar, además, que si el vocablo"*a"* hubiese aparecido en primer lugar de una lista de *16* términos, la probabilidad aleatoria de "ocurrencia" sería de *0.0625*, y la de "no-ocurrencia" *0.9375*[6]. Para una relación de *rango 20*, los valores habrían sido *0.05* y *0.95*, respectivamente. Este hecho aviva las dudas sobre la procedencia, o no, de otorgar en cualquier circunstancia un mismo *peso* a una misma *posición*, o su establecimiento atendiendo al *rango* de la lista más extensa de la *muestra* encuestada.

El modelo probabilístico se ha descrito en estas líneas en su aspecto más básico. La consideración de los individuos insertos en una colectividad de habla puede aconsejar tomar en cuenta otros parámetros: el total de vocablos utilizados por el conjunto de la comunidad, las longitudes de lista representativas o el número de casos en que se alcanza cada *posición*. La atención sobre esta clase de planteamiento nos arrastra a la explotación de los teoremas de la *probabilidad condicional* y a trabajar conceptos de índole *bayesiana*[7]. Tampoco es descabellado probar cocientes de *correlación* que valoren la probabilidad para que una palabra, conocida en un colectivo, sea usada por un sujeto junto a la posibilidad de su actualización en una determinada *posición*. En otros trabajos se profundiza en esta propuesta inicial, se analizan las soluciones apuntadas y se justifica la alternativa ensayada haciendo uso de los resultados de la investigación objeto de este trabajo.

BIBLIOGRAFÍA

AMÓN, Jesús. (1982). *Estadística para psicólogos (Probabilidad y estadística inferencial).* Tomo II. Neurociencia. Editorial Pirámide. Madrid.

---- (1990). *Estadística para psicólogos I. (Estadística descriptiva)*. Psicología, 12. Pirámide, Madrid.

ATO GARCÍA, Manuel. (2000). *Del contraste de hipótesis al modelo estadístico.* Grupo ModEst. Barcelona.

BAYES, Thomas. (1908). *Versuch zur Lösung eines Problems der Wahrscheinlichkeitsrechnung*. Engelmann. Leipzig.

BUTRÓN, Gloria Izaskun. (1987). *El léxico disponible: índices de disponibilidad*. Disertación presentada en la Universidad de Puerto Rico como requisito para la obtención del grado de Doctor en Filosofía y Letras. Puerto Rico.

[5] Cálculo de la probabilidad complementaria: 1 - 0,16 = 0,83.

[6] 1/16 = 0.625; 1-0.625 = 0.9375.

[7] La fórmula de Bayes [BAYES. 1908] es una consecuencia de la probabilidad condicional. Tiene en cuenta probabilidades *a priori* y probabilidades *a posteriori* y deriva en la denominada probabilidad bayesiana.

Web personal: http://angarmegia.tresuvesdobles.com - http://angarmegia.95mb.com
Correo electrónico: angarmegia@telefonica.net

CANAVOS, George C. (1988). *Probabilidad y estadística. Aplicaciones y métodos*. McGraw-Hill Interamericana de México. México.

CAO ABAD, Ricardo. (2001). *Introducción a la estadística y sus aplicaciones*. Pirámide. Madrid.

FERNÁNDEZ DE TROCONIS, Antonio. (1980). *Introducción a las teorías de las probabilidades. Estadística clásica y estadística bayesiana.* Bom. Bilbao.

FOX, David J. (1981). *El proceso de investigación en educación*. Ediciones Universidad de Navarra. S.A. Pamplona.

ITO, Kiyosi. (1984). *Introduction to probability theory*.Cambridge Univ. Press.Cambridge

GARCÍA MEGÍA, Antonio. (2004) *La disponibilidad léxica en la ciudad de Almería entre los nueve y los doce años.* Servicio de Publicaciones de la Universidad de Almería. Almería

GUILFORD, J. P. y FRUCHTER, Benjamín. (1984). *Estadística aplicada a la psicología y la educación*. McGraw-Hill Interamericana S.A. México.

KOHAN, N. y CARRÓ, J. (1968). *Estadística aplicada*. Eudeba. Buenos Aires.

LÓPEZ CHÁVEZ, Juan y STRASSBURGER FRÍAS, Carlos. (1987). *Otro cálculo del índice de disponibilidad léxica*, (comunicación). Noviembre de. 4ª Simposio de la Asociación Mexicana de Lingüística Aplicada. México.

SIEGEL, Sidney. (1983). *Estadística no paramétrica aplicada a las ciencias de la conducta*. 8ª ed.. Biblioteca Técnica de Psicología. Editorial Trillas, México.