

Media proporcional y frecuencia probabilística de las preguntas confirmativas españolas

Macronálisis y micronálisis de los datos de PRESEEA en el sistema LYNEAL

Hiroto Ueda, Universidad de Tokio

1. Introducción¹

Las formas de preguntas adicionales, por ejemplo, *¿no?*, *¿eh?*, *¿verdad?*, etc. que el hablante pone detrás de la afirmación para pedir la confirmación o simplemente llamar la atención del interlocutor son importantes tanto en su función comunicativa como en su uso práctico de alta frecuencia. Se supone que es uno de los fenómenos lingüísticos universales. Están estudiados en distintos campos de la lingüística en general y, en la lengua española en particular, contamos con estudios descriptivos del español coloquial y conversacional (Beinhauer 1978, Steel 1985, Briz 1998), la mayoría de los cuales se enfocan en sus características cualitativas. De todas las denominaciones ahí propuestas, adoptamos «pregunta confirmativa», puesto que es útil por ser fácilmente identificable con las formas en cuestión, a pesar de que en la práctica actual su función es frecuentemente control de contacto, como veremos más adelante.

El propósito del actual estudio es, tras revisar brevemente los estudios anteriores, investigar los aspectos cuantitativos de los usos reales documentados en los corpus de muestra del proyecto PRESEEA «Proyecto para el Estudio Sociolingüístico del Español de España y de América» (Moreno Fernández, 2005)². Propondremos los nuevos métodos de análisis numéricos, Media Proporcional y Frecuencia Probabilística, aplicables a los datos lingüísticos que presentan la distribución peculiar en forma de la letra L mayúscula. Finalmente, en la sección de Informática, presentaremos el sitio donde se puede comprobar las distintas funciones de LYNEAL «Letras y Números en Análisis Lingüísticos»³ con los datos concretos de PRESEEA⁴.

¹ Agradezco la ayuda generosamente prestada por Francisco Moreno Fernández, Ana María Cestero Mancera, Leyre Martín Aizpuru y Antonio Moreno Sandoval durante la realización de este estudio.

² Véase: <http://preseea.linguas.net/>

³ <http://lecture.ecc.u-tokyo.ac.jp/~cueda/lyneal/>

2. Estudios lingüísticos

2.1. Español coloquial

Los estudios lingüísticos y manuales de texto que tratan las expresiones coloquiales del español mencionan distintas formas para expresar las preguntas confirmativas. El manual del grupo de Bolinger (1973: 52-53) ofrece unas instrucciones útiles sobre las formas españolas, ¿eh?. ¿no?, ¿verdad?, en comparación con las inglesas.

- Por eso la invitación, ¿eh? / [sin comparación]
- Tiene uno portátil, ¿no? (¿verdad?)
/- She has a portable one, hasn't (doesn't) she?
- No regresa, ¿verdad? / - She isn't coming back, is she?

Es interesante la observación de los autores norteamericanos sobre el uso de ¿verdad?, que es, según ellos, correspondiente al 'tag' inglés de confirmación: *You don't like it, do you? You like it, don't you? He won't eat it, will he? He'll eat it, won't he?*, todos con el tono ascendente. Por otra parte, describen que ¿eh? es equivalente a *right?, huh?, hm?* Nos preguntamos por qué no han hecho corresponder ¿verdad? directamente a *right?*, que parece ser el equivalente léxico.

Beinhauer en su obra clásica (1978: 395-396) supone que la forma ¿verdad? es resultado de la reducción de la frase completa ¿No es verdad? (correspondiente al francés *n'est-ce pas*): *Esto era un convento, ¿verdad?* Por otra parte, piensa que en la pregunta ¿no?, se sobreentiende ¿(no) es verdad?: *Es muy serio, ¿no?* Deberíamos averiguar si estos dos procesos de reducción son históricamente comprobables o simplemente son unas reconstrucciones subjetivas e impresionistas. Nos llama la atención la observación que el investigador hace del uso especialmente frecuente de ¿no?, en sustitución de ¿verdad?, en las hablas americanas.

Steel (1985: 157-158) enumera las formas de preguntas que buscan (o asumen) el acuerdo positivo: ¿No es verdad?, ¿Verdad?, ¿Verdad que sí?, ¿No?, ¿Eh?, ¿No es así?, ¿No es eso?, ¿A que sí? y las compara con las formas inglesas equivalentes: *didn't he?, won't they? isn't she?, can't they?, right?, eh?*, etc. Afirma que las formas cortas ocurren con más frecuencia que las formas largas. Suponemos que la longitud tratada se refiere al número de palabras, pero también es posible pensar en su número total de

⁴ <http://lecture.ecc.u-tokyo.ac.jp/~cueda/lyneal/preseea.htm>

letras, fonemas o sílabas. Sus ejemplos son: *Vendrás pronto, ¿no?* / *Te gustan, ¿verdad?* / *No está mal la comida, ¿eh?*, (...) / *Crees que a la larga todo esto se arreglará, ¿no es así?* / *Las estrellas están en el aire, ¿no es eso?* / *Papá es muy bueno, ¿a que sí?*

2.2. Fonética

Dos obras de fonética de Quilis (1981, 1993) tratan sobre las configuraciones de entonación en las «preguntas confirmativas»⁵. En la primera obra (1981: 440) expone el siguiente patrón: / (1) 2 1 1 ↓ 1 2 ↑ / que caracteriza la pregunta confirmativa, tanto afirmativa como negativa, y agrega un aspecto fonético importante, falta de pausa entre los dos constituyentes de la oración (id.: 442): "Observando las curvas de configuraciones puede verse cómo en ninguno de los cuatro casos hay pausa: la terminación /-1 ↓/ desempeña aquí la función delimitadora (...)". Creemos que esta observación es importante porque es contraria a lo que observamos visualmente una coma en la parte indicada: *Es bonita* (↓), *¿verdad?* (↑) / *Es más grande* (↓), *¿no?* (↑).

Otra obra de fonética del mismo autor (1993: 451-452) presenta un comentario significativo: "La *pregunta confirmativa* se caracteriza por la presencia en el final del enunciado de una palabra prácticamente lexicalizada (*no*, *verdad*) con fundamental ascendente (...)." Debemos tratar con debida atención esta nota sobre la «lexicalización», es decir, no se trata de un uso más de algunas palabras comunes, sino más bien de unas palabras especialmente lexicalizadas e independientes.

2.3. Lexicografía

Los diccionarios, vocabularios o repertorios de expresiones orales del español presentan unos ejemplos usuales de preguntas confirmativas. Por ejemplo, el diccionario de Takahashi (1998: 247-248) colecciona los siguientes: *¿te enteras?*, *¿estamos?*, *¿lo oyes?*, *¿[me] entiendes?*, *¿entendido?*, *¿comprendido?*, *La próxima vez no te perdono. ¿Estamos?* / *El médico soy yo, ¿lo oyes?*, y *no usted*.

El repertorio de Gelabert et al. (1993: 129-130) clasifica las formas en tres grupos de acuerdo con el grado de aprendizaje: ELEMENTAL: *¿Me entiendes?*, *¿(Me) comprendes?*, *¿Comprendido?*, *¿Lo entiendes bien?* *¿Ya?*, *¿Vale?*, *¿De acuerdo?*. *¿Está claro?*, *¿Me explico?*, *¿Puedo continuar?*, *¿Podemos seguir?* / INTERMEDIO: *¿Lo ves?*, *¿Estás?*, *¿Estamos?*, *¿Me sigues?*, *¿Te das cuenta?* / AVANZADO: *¿Te aclaras?*, *¿Caes?*, *¿Lo captas?*

⁵ De ahí seguimos utilizando el término «pregunta confirmativa» presentado por Quilis (1981, 1993).

El proyecto lexicográfico de la Universidad de Augsburgo, dirigido por Haensch y Werner, ha venido publicando una serie de *Nuevo diccionario de americanismos*, dentro de la cual recurrimos al tomo de colombianismos para ver el uso peculiar de *¿cierto?*, empleado "para buscar el asentimiento de la persona que escucha, o como muletilla" y que corresponde al español de España, *¿verdad?* Otro ejemplo interesante es la forma característica de Uruguay *¿ta?*, *¿tamos?*, que describe Kühl de Mones (1993: s.v.): "*coloq.* Se usa al final de una oración interrogativa cuando se espera una respuesta afirmativa". También hay que anotar las formas conjugadas del verbo *ver*, *¿viste?*, *¿vio?*, *¿vieron?*, que se usan "a lo largo de una conversación como muletilla, sin ningún significado especial".

Por otra parte, el *Diccionario de americanismos* de la Asociación de Academias de la Lengua Española (2010), en la entrada de *¿cierto?* ofrece una información geográfica más amplia y detallada del uso en la actualidad: Colombia y oeste de Bolivia.

2.4. Marcador discursivo

Dentro de las distintas formas del marcador discursivo en general, que es una expresión que muestra la conexión entre lo que se está diciendo y el contexto o indica la actitud del hablante a lo que está diciendo, las preguntas confirmativas constituyen un subgrupo de «marcador conversacional». Los estudios del mismo marcador coinciden en afirmar su función de mantenimiento de contacto con el interlocutor. En efecto, Casado Velarde (1998: 65) las clasifica en el apartado de «mantenimiento de atención interlocutiva» y registra las siguientes: *¿no?*, *¿verdad?*, *¿no sabes?*, *¿ves?*, *¿oyes?*, *¿eh?*, *¿comprendes?* Baixauli Fortea (2000: 96) habla del "control de la atención y comprensión del mensaje" con un ejemplo de: [...] *venimos a arreglarlo (...) ¿sabes? así que allí está el reloj en mi casa*, y afirma que su finalidad es expresiva, apelativa y fática.

En los «apéndices con valor apelativo» de Fuentes Rodríguez (1990: 172) "hay dos grupos claramente diferenciados por su forma y su funcionamiento": uno en forma imperativa (*fíjate*, *imagínate*, *oye*, *mira*) y otro en forma interrogativa (*¿verdad?*, *¿no?*, *¿sabes?*, *¿entiendes?*, *comprendes?*, *¿ves?*). En cuanto al segundo grupo, el tema del presente estudio, explica de la forma *¿verdad?*, tanto intercalada en el enunciado, como al final del mismo: "Su función es servir de apoyo en la conversación, apelando al oyente para que siga así, escuchando al hablante. Ha perdido (...) el valor de solicitar una comprobación a este. Su uso es fático." Lo mismo ocurre con la forma *¿no?*, con la

que, según Ortega (198: 247), "el hablante crea al oyente la obligación de responder y espera que se confirmen sus previsiones". Sin embargo, para Fuentes Rodríguez (ibid.: 185), esto era así solo en el origen, y piensa que los datos muestran "cómo esa conciencia se va debilitando y se pierde esa obligación de responder, esa necesidad. El hablante (...) apela al oyente, intenta mantener el contacto con él, se justifica en cierto modo." También de la forma *¿sabes?*, explica (ibid.: 188): "No es que el hablante justifique el hecho de hablar porque es importante que el oyente se entere de esto, sino que con *¿sabes?* quiere apelar al oyente para que actúe como al oyente e integre en sus conocimientos el que acaba de dársele. Llama la atención del oyente sobre lo dicho". Deberíamos comprobar en la realidad de la conversación si de hecho los hablantes esperan la respuesta del interlocutor⁶.

Briz (1998) estudia estos marcadores discursivos de control del contacto en relación con su posición con respecto al resto del enunciado. Respecto al tonema ascendente observa que "la posición final en el enunciado de estos marcadores así como su marcado ascenso final, contrasta con la posición interior y el tonema final «escasamente ascendente» de aquellos otros, simples continuadores fáticos." Sus ejemplos son: *Le he dado ¿sabes? muchas vueltas a lo que me dijiste el otro día / ¿No estabas de régimen? / pues los dulces engordan ¿sabes?* (ibid.: 225) Y en cuanto a la interpretación compara la posición final absoluta en una intervención que "favorece la interpretación reafirmativa-exhortativa (función apelativa)" y otras posiciones, donde se realiza la "simple reafirmación (función-fática)" (ibid.: 226).

La observación de la función fática en estos elementos la corrobora también Cortes Rodríguez y Camacho Adarve (2005: 174-176):

Los del grupo siguiente (*¿me explico?*, *¿verdad?*, *¿comprende?*) son mecanismos de comprobación o justificación (...) de que el oyente está inmerso de interesado en la interacción; (...) Esta clase de marcadores son nuestros *marcadores de cierre*,

⁶ Efectivamente en *El arrepentido* de Ana María Matute (1967), encontramos dos ocurrencias de la forma *¿sabes?*, con la que el hablante no parece esperar necesariamente la respuesta del otro: "'- Está bien, Ruti. Te lo agradezco, ¿sabes?... Sí; te lo agradezco mucho. Es mejor así. / Ruti guardó silencio. (...) // - (...) tengo mucho dinero, Ruti. ¿sabes? No siempre las cosas son como parecen.' / Ruti sonrió." El traductor de la misma obra no las interpreta ni como preguntas: "'Fair enough, Ruti. I'm grateful to you, *you know*... Yes, I'm very grateful to you. It's better that way.' / Ruti kept quiet. (...) // '(...) I have lots of money, Ruti. Things aren't always what they seem, *you know*.' / Ruti smiled." Lawaetz (1972: 10-13). Se nota que Tomeu el Viejo, el que habla, busca desesperadamente el contacto sentimental con Ruti, su sobrino.

llamados en la bibliografía internacional «tag question», de turno, cuyas formas pueden ser justificativas (*¿me comprendes?, ¿ves?, ¿sabes?* o de comprobación *¿eh?, ¿verdad?*) (...) Nosotros añadimos que si la entonación es imperativa, automáticamente se convierten en antipáticos, o de alejamiento, como *¿queda claro?*

El estudio de Cestero Mancera (2000: 618), basado en su propio corpus de conversaciones entre dos personas y discursos académicos, subraya el aspecto colaborativo de la conversación⁷:

(...) tanto el emisor como el receptor han de participar de forma activa en la construcción de la interacción y, por tanto, han de asegurar la apertura continua del contacto: a) Enunciados producidos por el emisor, esto es, recursos utilizados por el hablante en su turno de habla para asegurarse de que el canal está abierto y el interlocutor sigue la comunicación. Ejemplos típicos son las preguntas veritativas o comprobativas del tipo de *¿no?, ¿verdad?, ¿comprendes?, ¿me sigues?* ... b) Enunciados con función fática producidos por el interlocutor, es decir, "apoyos" emitidos por el oyente para asegurar al hablante su seguimiento puntual y la apertura continua del canal. Ejemplos típicos son *sí, claro, es verdad*.

3. Tratamientos estadísticos

3.1. Cuestiones del estudio

Tras revisar los estudios anteriores sobre las formas de la «pregunta confirmativa», vamos a investigar las cuatro siguientes cuestiones al respecto:

- ¿Qué formas existen en el inventario de las preguntas confirmativas en un corpus estrictamente estructurado y suficientemente amplio?
- ¿Cuál es la realidad numérica, en términos de distribución de frecuencia, de las preguntas confirmativas observadas en el corpus?
- ¿Qué importancia cuantitativa tiene cada forma de pregunta confirmativa en el corpus?
- ¿Cómo se presenta la variación según los parámetros extralingüísticos en el

⁷ Para los detalles de estas funciones y cada forma concreta, véase también Cestero Mancera (2003).

corpus?

Para estos objetivos hemos utilizado el corpus PRESEEA por su característica apropiada para estudiar la conversación registrada en distintos lugares de habla española, con atributos controlados con los mismos criterios extralingüísticos: lugar, sexo, edad y nivel de educación. De esta manera estamos en condiciones de analizar los datos relevantes con rigor estadístico.

3.2. Descripción de frecuencias

Hemos encontrado 28 formas siguientes (total: 7262)⁸: *¿no?* (frecuencia: 4608), *¿sí?* (874), *¿eh?* (467), *¿cierto?* (341), *¿verdad?* (235), *¿cachái(s)?* (173), *¿sabes?* (158), *¿viste?* (116), *¿hm?* (101), *¿ah?* (57), *¿entiendes?* (27), *¿ves?* (16), *¿as?* (14), *¿te acuerdas?* (9), *¿comprendes?* (9), *¿te parece?* (8), *¿oíste?* (7), *¿vale?* (5), *¿no es cierto?* (5), *¿no es verdad?* (5), *¿no le parece?* (5), *¿ta?* (5), *¿correcto?* (4), *¿te das cuenta?* (4), *¿de acuerdo?* (3), *¿estamos?* (2), *¿no cree?* (2), *¿mirá?* (1), *¿ok?* (1). Se nota la distribución sumamente sesgada, es decir, existen unas pocas formas extremadamente frecuentes, mientras que la mayoría ocurre menos de 10 veces dentro de la totalidad. Y dentro de las formas frecuentes tampoco observamos una gradación descendente suave sino que el descenso es tan abrupto como de manera exponencial negativa. Si dibujamos un gráfico de estas frecuencias, la línea presentará la forma de la letra L mayúscula, como veremos inmediatamente.

Pero antes veamos la lista de las preguntas confirmativas dentro de la totalidad de las preguntas monoléxica, confirmativas o no, de las cuales destacamos las preguntas confirmativas con asterisco (*) en negrita:

Cuadro 3.1. Pregunta monoléxica sin agrupación

R	Forma	Suma	Cumul.	%
1	*¿no?	4608	4608	55.1
2	*¿sí?	874	5482	65.5

⁸ En la lista de las preguntas confirmativas, hemos agrupado las variantes en una forma, por ejemplo, *¿eh?* = {*¿e?*, *¿eh?*, *¿eeh?*} Hemos utilizado las expresiones regulares siguientes: *¿no?*, *¿sí?*, *¿e+h?*, *¿cierto?*, *¿verdad?*, *¿cachái(s)?*, *¿sab(é|e)s?*, *¿viste?*, *¿(hm|hum|mhm|uhum|m+)?* con añadidira de "<.*?>=>", con la que se excluyen las etiquetas puestas entre paréntesis angulares "<...>"

3	<i>*¿eh?</i>	453	5935	71.0
4	<i>*¿cierto?</i>	341	6276	75.0
5	<i>*¿verdad?</i>	277	6553	78.3
6	<i>*¿cachái(s)?</i>	163	6716	80.3
7	<i>*¿sabes?</i>	151	6867	82.1
8	<i>¿qué?</i>	128	6995	83.6
9	<i>*¿viste?</i>	116	7111	85.0
10	<i>¿ya?</i>	75	7186	85.9
11	<i>¿cómo?</i>	65	7251	86.7
12	<i>*¿hm?</i>	61	7312	87.4
13	<i>*¿entiendes?</i>	58	7370	88.1
14	<i>*¿ah?</i>	55	7425	88.8
15	<i>*¿mmm?</i>	21	7446	89.0
16	<i>¿dónde?</i>	20	7466	89.3
17	<i>¿yo?</i>	20	7486	89.5
18	<i>¿cuál?</i>	18	7504	89.7
19	<i>*¿sí?</i>	15	7519	89.9
20	<i>¿también?</i>	15	7534	90.1

La tabla reproduce las preguntas monolécicas hasta el número 20 de rango, en orden descendiente. Dentro de las 10 formas monolécicas más frecuentes, la mayoría (9) son preguntas confirmativas y dentro de las 20 formas más frecuentes, más de la mitad (12) son preguntas confirmativas. Esta lista demuestra el carácter principalmente monolécico de la pregunta confirmativa y afirma la observación de Steel (1985: 158), que hemos visto en 2.1. De nuestra parte, proponemos caracterizar la pregunta confirmativa por su rasgo monolécico, más bien que por la longitud de la forma medida por el número de letras, fonemas o sílabas, puesto que dentro de las formas enumeradas en la lista anterior, encontramos formas de distintos números de fonemas y sílabas.

A partir de la lista anterior dibujamos el siguiente gráfico de línea:

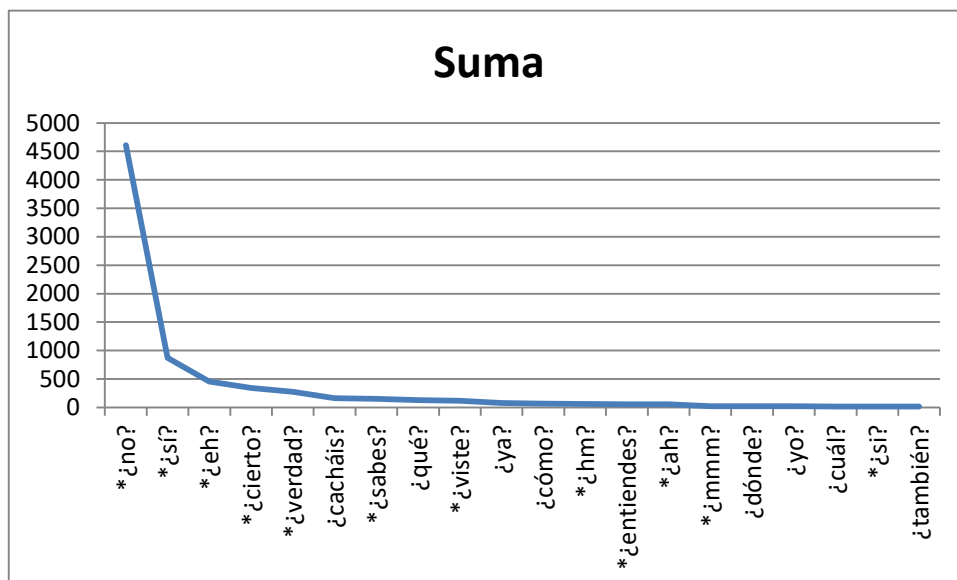


Fig. 3.1. Suma de las preguntas monoléxicas

En esta figura observamos el descenso pronunciado de la frecuencia en las primeras formas más altas, lo que dibuja la figura de la letra L. Parece ser que las preguntas confirmativas están tan lexicalizadas que no permiten cambios paradigmáticos con otras palabras, a diferencia de las formas no lexicalizadas como, por ejemplo, *¿ahora?*, *¿mañana?*, *¿veinte?*, lo que comprueba la observación de Quilis (1993: 451-452, véase 2.2).

La diferencia de esta lista con respecto a los estudios anteriores (véase 2) está en que incluye las formas locales (*¿cierto?*, *¿cachái(s)?*, *¿viste?*) y presenta un orden de frecuencia que es difícil de saber por la observación impresionista basada en las experiencias personales.

3.3. Macroanálisis

En esta sección veremos las frecuencias en relación con los atributos que poseen los informantes⁹: lugar, sexo, edad y nivel de educación. Este análisis común y usual en investigaciones lingüísticas, lo denominamos «Macroanálisis», puesto que se analiza la totalidad con división por categorías preestablecidas de atributos. En la

⁹ El corpus de muestras de PRESEEA está constuido de 18 textos correspondientes a 18 informantes, que se clasifican por 2 sexos, multiplicado por 3 grupos de edad, multiplicado por 3 niveles de educación, es decir 18 informantes. En esta sección 3.3. y en la siguiente 3.4, tratamos los textos recogidos de I (Informante) y E (Encuestador(a)), a exclusión de otros participantes minoritarios.

sección siguiente, en cambio, intentamos realizar otro análisis, denominado «Microanálisis», donde se analizan personas en combinación con atributos no divididos.

(1) LUGAR

La tabla siguiente muestra la Frecuencia Probabilística (FP)¹⁰ multiplicada por 100000 de las distintas preguntas confirmativas del español en las nueve ciudades hispanohablantes, con añadidura de la Media Proporcional¹¹ (MP), que representa la magnitud media del conjunto¹²:

Cuadro 3.3.1. Pregunta confirmativa / lugar

FP.:100000	Es.A	Es.M	Es.V	Cu.H	Mx.M	Co.M	Pe.L	Ch.S	Ur.M
1. ¿no?	221.8	591.1	180.0	190.3	94.0	6.5	1382.5	22.8	272.5
2. ¿sí?	84.2	181.0	8.5	0.3	49.3	12.7	3.6	20.8	95.8
3. ¿eh?	52.9	79.4	84.1	2.3	10.4	0	0.1	0	9.8
4. ¿cierto?	0	0	0	0	0	210.0	0	5.7	0
5. ¿verdad?	2.4	8.5	0.1	1	58.5	0	0	0	4.9
6. ¿cachái(s)?	0	0	0	0	0	0	0	78.1	0
7. ¿sabes?	24.7	40.6	2.0	0	0	0	0	0	0
8. ¿viste?	0	0	0	0	0	0	0	0.6	63.3
9. ¿hm?	0	27.2	0	0	0	0	0	6.1	0.2
MP	154.7	421.3	144.0	186.8	69.7	193.3	1378.8	53.1	196.1

Nos llama la atención la frecuencia sumamente alta de *¿no?* en Lima (Perú), lo que confirmaría parcialmente la observación de Beinhauer (1978: 396), quien afirmaba la alta frecuencia general de los hablantes americanos en comparación con los hablantes de España. Notamos, sin embargo, que la frecuencia más alta se concentra en Lima, a la que sigue Madrid (España) y Montevideo, mientras que otros lugares de Hispanoamérica no presentan cifras importantes.

Por otra parte, comprobamos localismos de preguntas confirmativas: *¿cierto?* en Colombia (Medellín), *¿cachái(s)?* en Chile (Santiago) y *¿viste?* en Uruguay (Montevideo). Curiosamente, la forma de la pregunta confirmativa *¿sabes?*, correspondiente a la inglesa usual *you know?*, no se ha recogido más que en España

¹⁰ Para la Frecuencia Probabilística, véase ADDENDA-2.

¹¹ Para la Media Proporcional, véase ADDENDA-1.

¹² Las ciudades encuestadas son: Es.A. (España, Alcalá), Es.M. (España, Madrid), Es.V. (España, Valencia), Cu.H. (Cuba, La Habana), Mx.M. (México, Monterrey), Co.M. (Colombia, Medellín), Pe.L. (Perú, Lima), Ch.S. (Chile, Santiago), Ur.M. (Uruguay, Montevideo).

(Alcalá, Madrid, Valencia). Consideramos que estas formas marcan destacadamente sus rasgos diatópicos caracterizando el modo de hablar de cada sitio.

Estos hechos locales son explicables desde el punto de vista de su carácter fático en la interacción (Fuentes Rodríguez 1990: 171-172, Briz 1998: 224-225, Casado Velarde, 1998: 65; Cortes Rodríguez / Camacho Adarve, 2005: 174-176; Cestero Mancera 2000, 618-621). No se trata de los vocablos descriptivos generales ni de palabras gramaticales comunes, sino de las expresiones que fomentan las relaciones humanas y, por esta razón, es preferible utilizar los localismos dialectales con los que se identifican como miembros de la misma comunidad.

(2) SEXO

En cuanto a la diferencia por sexo, no se observa gran diferencia entre ellos, tanto en las formas individuales, como el conjunto representado por la Media Proporcional (MP). Tan solo destacamos la frecuencia relativamente alta de *¿cierto?* en Hombres y *¿cachái(s)?* en Mujeres. Sin embargo, en la sección anterior hemos anotado que las dos formas son localizables en Colombia y en Chile, respectivamente. Por lo tanto, deberíamos indagar sus detalles en cada localidad, como veremos en 3.4.

Cuadro 3.3.2. Pregunta confirmativa / sexo

FP.Pl.:100000	Hombre	Mujer	(No se sabe)
1. ¿no?	375.9	284.7	248.7
2. ¿sí?	43.2	73.9	85.4
3. ¿eh?	34.4	23.9	0
4. ¿cierto?	41.5	1.9	0
5. ¿verdad?	9	17.1	0.1
6. ¿cachái(s)?	0.5	18.1	0
7. ¿sabes?	4.5	11.9	0.1
8. ¿viste?	7.1	5.0	0
9. ¿hm?	2.2	6.9	0
MP	282.2	198.3	206.9

(3) EDAD¹³

Con respecto a la EDAD tampoco observamos diferencias destacables entre los tres grupos en la Media Proporcional (MP). En cada forma, sin embargo, se notan

¹³ EDAD: Edad-1:20-34; Edad-2: 35-54; Edad-3: 55-.

ciertas concentraciones numéricas interesantes. Lo mismo que el caso anterior, estas afirmaciones deberíamos comprobarlas al detalle en la sección siguiente (3.4), donde tratamos de realizar el microanálisis de los casos concretos.

Cuadro 3.3.3. Pregunta confirmativa / edad

FP.Pl.:100000	Edad-1	Edad-2	Edad-3	(No se sabe)
1. ¿no?	356.9	392.5	199.7	249
2. ¿sí?	60	90.3	7.9	229
3. ¿eh?	6.1	35.3	47.2	0.2
4. ¿cierto?	3.4	52.4	1.7	0
5. ¿verdad?	18.4	11.6	5.3	1.4
6. ¿cachái(s)?	18.4	8.7	0	0
7. ¿sabes?	10.0	11.1	0.7	3.6
8. ¿viste?	6.8	4	3.1	0
9. ¿hm?	0.4	9.1	3.1	0
MP	274.4	271	157.2	236.9

(4) NIVEL DE EDUCACIÓN¹⁴

Según la tabla siguiente, las personas del nivel bajo de educación no utilizan las preguntas confirmativas tanto como los de los niveles superiores de educación. Esta observación se precisa por la división entre Nivel-2 y Nivel-3 en cada caso de la forma. Veamos los detalles en el Microanálisis en la sección siguiente (3.4).

¹⁴ NIVEL DE EDUCACIÓN: Nivel-1: Analfabeto, enseñanza primaria, 5 años de escolarización; Nivel-2: Enseñanza secundaria, 10-12 años de escolarización, Nivel-3: Enseñanza superior (universitaria, técnica superior), 15 años de escolarización.

Cuadro 3.3.4. Pregunta confirmativa / nivel de educación

FP.Pl.:100000	Nivel-1	Nivel-2	Nivel-3	(No se sabe)
1. ¿no?	163.3	381.8	374.6	247.8
2. ¿sí?	1.5	11.7	118.8	227.9
3. ¿eh?	37.2	19.8	28.5	0.2
4. ¿cierto?	1.6	0.1	45.8	0
5. ¿verdad?	3.4	11.1	18.3	1.3
6. ¿cachái(s)?	4.2	14.8	7	0
7. ¿sabes?	8.6	12.3	2.7	3.5
8. ¿viste?	4.8	12.5	0.4	0
9. ¿hm?	0	1.2	9.7	0
MP	125.5	315.8	260.5	235.8

3.4. Microanálisis

En esta sección nos dedicamos a analizar los datos con respecto a la combinación de atributos. Mediante la combinación de atributos se identifica el individuo, de modo que en lugar de ver las frecuencias en las categorías preestablecidas (por ejemplo, distintos lugares investigados o el sexo), observaremos la frecuencia de uso de la forma interrogativa en cuestión registrada en cada persona identificada con la combinación de atributos: lugar, sexo (hombre / mujer), edad (1/2/3), nivel de educación (1/2/3) y relación (conocido / desconocido). En otras palabras, el propósito del análisis individual consiste en estudiar los datos a partir de cada caso (forma lingüística) en cada persona (hablante) sin criterios preestablecidos para llegar a la categorización *a posteriori*. Este método lo denominamos «Microanálisis», por tratarse de cada caso individual en el inicio.

(1) ¿no?

Como hemos visto anteriormente, la forma más frecuente de la pregunta confirmativa es *¿no?*. Ahora veamos su distribución por los hablantes concretos. La tabla siguiente está constituida por R (rango), Hablante, FP (Frecuencia Probabilística), Acum. (Acumulación de FP) y % (porcentaje de la Acumulación con respecto a la totalidad). Nos interesa sobremanera las primeras personas que utilizan la forma con alta frecuencia. A modo del Análisis ABC (véase leanManufacturing en la Referencia), delimitamos la lista en las primeras 32 personas, con las que el porcentaje acumulado sobrepasa el 80%. Esto quiere decir que fijándonos en estas primeras 32 personas,

cubrimos la inmensa mayoría (más de 80%) de los usos de ¿no? en función confirmativa.

Cuadro 3.4.1.a. ¿no? / atributos del hablante

R	¿no?: Hablante	FP	Acum.	%
1	PE-LIM+I+H+1+3	383.0	383.0	9.0
2	ES-MAD+I+H+2+2	241.5	624.5	14.6
3	PE-LIM+I+H+2+2	224.0	848.5	19.8
4	PE-LIM+I+M+2+2	192.5	1041.0	24.3
5	PE-LIM+I+H+2+1	189.4	1230.4	28.8
6	PE-LIM+I+M+1+3	163.5	1393.9	32.6
7	ES-MAD+I+H+2+3	148.9	1542.8	36.1
8	PE-LIM+I+H+1+2	146.1	1688.9	39.5
9	PE-LIM+I+H+3+2	133.4	1822.3	42.6
10	PE-LIM+I+M+3+3	115.0	1937.3	45.3
11	ES-VAL+I+H+1+2	110.0	2047.3	47.9
12	CU-HAB+I+H+2+2	106.0	2153.3	50.4
13	ES-ALC+E+M+1+3	91.6	2244.9	52.5
14	PE-LIM+E+M+1+3	87.9	2332.8	54.6
15	ES-MAD+E+M+2+3	87.6	2420.4	56.6
16	PE-LIM+E+M+1+3	85.8	2506.2	58.6
17	PE-LIM+I+M+2+3	85.3	2591.5	60.6
18	PE-LIM+I+H+3+3	79.7	2671.2	62.5
19	ES-MAD+I+M+1+2	74.4	2745.6	64.2
20	ES-MAD+I+H+1+2	73.0	2818.6	65.9
21	UR-MTV+E+M+2+3	72.7	2891.3	67.6
22	ES-MAD+E+H+2+3	69.1	2960.4	69.2
23	PE-LIM+I+M+1+2	67.9	3028.3	70.8
24	PE-LIM+I+M+3+2	65.3	3093.6	72.3
25	ES-MAD+I+H+2+1	62.6	3156.2	73.8
26	UR-MTV+I+M+3+3	61.9	3218.1	75.3
27	ES-MAD+I+H+3+2	39.1	3257.2	76.2
28	UR-MTV+E+H+3+3	37.4	3294.6	77.0
29	UR-MTV+I+H+1+3	37.1	3331.7	77.9
30	PE-LIM+I+M+2+1	35.2	3366.9	78.7
31	ES-ALC+I+M+2+2	35.1	3402.0	79.6

Dentro de los múltiples parámetros, el LUGAR es destacable, seguido de EDAD y NIVEL DE EDUCACIÓN, de modo que cruzamos LUGAR con EDAD y NIVEL DE EDUCACIÓN. Ahora es conveniente enfocarnos en la distribución de frecuencias del conjunto mayoritaria de manera cruzada:

Cuadro 3.4.1.b. ¿no?: LUGAR / EDAD

LUGAR	Edad-1	Edad-2	Edad-3	Suma
PE-LIM	934.2	726.4	393.4	2054.0
ES-MAD	147.4	609.7	39.1	796.2
UR-MTV	37.1	72.7	99.3	209.1
ES-ALC	91.6	35.1		126.7
ES-VAL	110.0			110.0
CU-HAB		106.0		106.0
MX-MON		31.6		31.6
Total	1320.3	1581.5	531.8	3433.6

Cuadro 3.4.1.c. ¿no?: LUGAR / NIVEL DE EDUCACIÓN

LUGAR	Nivel-1	Nivel-2	Nivel-3	Suma
PE-LIM	224.6	829.2	1000.2	2054.0
ES-MAD	62.6	428.0	305.6	796.2
UR-MTV			209.1	209.1
ES-ALC		35.1	91.6	126.7
ES-VAL		110.0		110.0
CU-HAB		106.0		106.0
MX-MON			31.6	31.6
Total	287.2	1508.3	1638.1	3433.6

De esta manera, podemos explicar la mayoría de los casos (80%) destacando la concentración en Lima, seguida de Madrid, preferiblemente en los dos grupos de EDAD jóvenes (Edad-1, Edad-2) y NIVEL DE EDUCACIÓN más alto (NIVEL-2, NIVEL-3). La observación de Beinhauer (1978: 396), quien menciona la alta frecuencia de ¿no? en las hablas americanas en general (véase 3.1), se precisa puntualmente en Lima, a exclusión de otras ciudades investigadas (Monterrey, Medellín, Santiago). Es destacable el uso frecuente en Madrid, seguida de Montevideo. Repetimos apuntar su distribución preferencial en edades menores y niveles de educación altos.

(2) ¿sí?

La segunda forma más frecuente ¿sí? llega fácilmente a más de 80% con solo 11 personas. Esta vez con tan solo 11 personas cubrimos más de 80% de la totalidad:

Cuadro 3.4.2.a. ¿sí? / atributos del hablante

R	¿sí?: Hablante	FP	Acum.	%
1	ES-MAD+E+M+1+3	129.9	129.9	17.0
2	ES-ALC+E+M+1+3	77.1	207.0	27.1
3	ES-MAD+E+H+2+3	73.5	280.5	36.7
4	ES-MAD+E+M+2+3	71.2	351.7	46.0
5	ES-MAD+E+M+2+3	70.0	421.7	55.2
6	UR-MTV+E+M+2+3	63.3	485.0	63.5
7	UR-MTV+E+M+2+3	41.5	526.5	68.9
8	ES-MAD+E+H+2+3	31.5	558.0	73.0
9	MX-MON+E+M+1+3	27.4	585.4	76.6
10	ES-ALC+E+M+x+x	25.0	610.4	79.9
11	UR-MTV+E+M+1+3	21.1	631.5	82.6

Hemos detectado los sesgos importantes en la tabla cruzada de LUGAR / SEXO y la de LUGAR / NIVEL.

Cuadro 3.4.2.b. ¿sí?: LUGAR / SEXO

LUGAR	Hombre	Mujer	Suma
ES-MAD	105.0	271.1	376.1
UR-MTV		125.9	125.9
ES-ALC		102.1	102.1
MX-MON		27.4	27.4
Total	105.0	526.5	631.5

Cuadro 3.4.2.c. ¿sí?: LUGAR / NIVEL

LUGAR	Nivel-3	(No se sabe)	Suma
ES-MAD	231.4	144.7	376.1
UR-MTV	84.4	41.5	125.9
ES-ALC		102.1	102.1
MX-MON	27.4		27.4
Total	606.5	25	631.5

La pregunta confirmativa con ¿sí? se realiza mayoritariamente en Madrid, Montevideo y Alcalá. Dentro de estas ciudades, con añadidura de la siguiente Monterrey, sus usuarias mayoritarias son mujeres con Nivel alto de educación.

(3) ¿eh?

La forma confirmativa ¿eh? cuenta con 15 personas para pasar el umbral de 80%. Las tablas cruzadas se han preparado en relación con EDAD y RELACIÓN entre el encuestador y el informante, más destacables que otros atributos:

Cuadro 3.4.3.a. ¿eh? / atributos del hablante

R	¿eh?: Hablante	FP	Acum.	%
1	ES-VAL+I+H+3+1	31.9	31.9	18.6
2	ES-MAD+I+M+2+2	17.3	49.2	28.7
3	ES-MAD+I+H+2+1	14.9	64.1	37.4
4	ES-ALC+I+H+1+3	11.0	75.1	43.8
5	ES-ALC+E+H+2+1	9.6	84.7	49.4
6	ES-VAL+I+M+2+3	8.8	93.5	54.6
7	ES-MAD+I+H+3+3	8.7	102.2	59.6
8	ES-MAD+E+M+2+3	7.6	109.8	64.1
9	ES-ALC+I+H+2+2	5.1	114.9	67.0
10	ES-MAD+I+M+3+3	4.7	119.6	69.8
11	ES-MAD+I+H+3+2	4.1	123.7	72.2
12	ES-VAL+I+M+3+1	4.1	127.8	74.6
13	ES-ALC+I+M+3+2	3.9	131.7	76.8
14	ES-VAL+I+M+3+3	3.5	135.2	78.9
15	UR-MTV+I+H+3+3	3.1	138.3	80.7

Cuadro 3.4.3.b. ¿eh?: LUGAR / EDAD

LUGAR	Edad-1	Edad-2	Edad-3	Suma
ES-MAD		39.8	17.5	57.3
ES-VAL		8.8	39.5	48.3
ES-ALC	11.0	14.7	3.9	29.6
UR-MTV			3.1	3.1
Total	11.0	63.3	64.0	138.3

Cuadro 3.4.3.c. ¿eh?: LUGAR / RELACIÓN

LUGAR	Conocidos	Desconocidos	Suma
ES-MAD	16.4	40.9	57.3
ES-VAL	3.5	44.8	48.3
ES-ALC		29.6	29.6
UR-MTV	3.1		3.1
Total	23.0	115.3	138.3

Por las dos tablas precedentes, afirmamos que la forma *¿eh?* es preferida por los hablantes de las tres ciudades españolas, Madrid, Valencia y Alcalá, a las que sigue, con gran distancia, Montevideo. Es utilizada por las personas de edades relativamente mayores (Edad-2 y Edad-3) en relación encuestador - informante de desconocidos.

(4) ¿cierto?

Dentro de las ciudades encuestadas, la forma *¿cierto?* es exclusiva de Medellín¹⁵. Efectivamente, la tabla siguiente muestra su uso predominante en Colombia. Como se trata de una forma casi exclusiva de una persona (CO-MED + I + H + 2 + 3), no hace falta ver sus detalles interiores para buscar tendencias destacables. PRESEEA la registra, aunque de manera minoritaria, también en Santiago de Chile.

Cuadro 3.4.4. ¿cierto? / atributos del hablante

R	<i>¿cierto?:</i> Hablante	FP	Acum.	%
1	CO-MED+I+H+2+3	129.1	129.1	89.4
2	CO-MED+I+H+3+1	5.0	134.1	92.9
3	CO-MED+I+H+1+3	2.0	136.1	94.3
4	CO-MED+I+M+3+3	1.9	138.0	95.6
5	CO-MED+I+M+2+3	1.8	139.8	96.8
6	CH-STG+19+H+1+3	1.7	141.5	98.0

(5) ¿verdad?

La forma confirmativa usual *¿verdad?* muestra su presencia predominante en Monterrey (México) entre las 9 ciudades investigadas. También es cierto que se registra en las tres ciudades españolas (Alcalá, Madrid, Valencia) y en otra cubana, La Habana

¹⁵ Hemos visto que Asociación de Academias de la Lengua Española (2010) registra *¿cierto?* no solamente en Colombia, sino en la parte oeste de Bolivia. Véase 2.3.

(véase 3.3). Sin embargo, se oscurece su importancia numérica en el aluvión de uso que viene de Monterrey.

Cuadro 3.4.5. ¿verdad? / atributos del hablante

R	¿verdad?: Hablante	FP	Acum.	%
1	MX-MON+E+M+1+3	28.6	28.6	30.2
2	MX-MON+E+H+1+2	10.8	39.4	41.6
3	MX-MON+E+M+2+2	10.2	49.6	52.4
4	MX-MON+E+M+1+2	9.6	59.2	62.6
5	MX-MON+E+M+2+3	6.6	65.8	69.6
6	MX-MON+E+H+1+2	5.7	71.5	75.6
7	MX-MON+E+M+1+3	4.9	76.4	80.8

(6) ¿cachái(s)?

La forma peculiar de Chile (Santiago), *¿cachái(s)?*, se confirma en la tabla siguiente. Las tres personas con las que llega a 87.2% de uso son mujeres:

Cuadro 3.4.6. ¿cachái(s)? / atributos del hablante

R	¿cachái(s)? Hablante	FP	Acum.	%
1	CH-STG+43+M+1+2	54.2	54.2	50.0
2	CH-STG+91+M+2+3	29.2	83.4	76.9
3	CH-STG+79+M+1+3	11.1	94.5	87.2

(7) ¿sabes?

El uso exclusivo de *¿sabes?* en España (véase 3.3) se confirma en las dos ciudades, Madrid y Alcalá, que son cercanas geográficamente:

Cuadro 3.4.7. ¿sabes? / atributos del hablante

R	¿sabes?: Hablante	FP	Acum.	%
1	ES-MAD+I+M+2+2	27.9	27.9	44.1
2	ES-ALC+I+H+1+3	11.0	38.9	61.5
3	ES-MAD+I+M+1+1	5.1	44.0	69.5
4	ES-ALC+I+H+2+2	3.8	47.8	75.5
5	ES-ALC+I+M+1+3	3.7	51.5	81.4

(8) ¿viste?

Como hemos visto anteriormente (2.3), *¿viste?* con función discursiva de pregunta confirmativa es una forma peculiar de Uruguay (Kühl de Mones, 1993: s.v. *ver*). Lo comprobamos con los detalles individuales en la tabla siguiente:

Cuadro 3.4.8. ¿viste? / atributos del hablante

R	¿viste?: Hablante	FP	Acum.	%
1	UR-MTV+I+H+1+2	43.7	43.7	58.3
2	UR-MTV+I+M+3+1	13.2	56.9	75.9
3	UR-MTV+I+H+2+2	9.8	66.7	88.9

(9) ¿hm?

La forma *¿hm?*, la novena en el ranking de frecuencia, no es específica de español, sino que también se utiliza en otras lenguas europeas. Incluso de vez en cuando la misma forma se escucha en la conversación japonesa, de modo que puede ser un fenómeno casi común en lenguas que no tienen relación genética. Suponemos que se caracteriza de manera no lingüística, sino más bien de manera paralingüística, de la misma manera que la risa, llanto, suspiro, carraspeo, bostezo, etc. que se transcriben con alfabetos más o menos cercanos: *ja ja, je je, ejem*, etc., que son también comunes entre las distintas lenguas. Efectivamente, encontramos varias transcripciones en el corpus PRESEEA¹⁶: *¿hm?*, *¿hum?*, *¿mhm?*, *¿uhm?*, *¿m?*, *¿mm?*. Por esta razón, la distribución siguiente no mostraría más que unos ejemplos registrados en Madrid y Santiago de Chile. Suponemos que se encontrarán más casos en nuestras investigaciones futuras con más datos transcritos.

Cuadro 3.4.9. ¿hm? / atributos del hablante

R	¿hm?: Hablante	FP	Acum.	%
1	ES-MAD+E+M+2+3	19.7	19.7	38.1
2	CH-STG+103+M+3+3	14.6	34.3	66.3
3	ES-MAD+E+H+2+3	5.6	39.9	77.2
4	ES-MAD+I+M+2+2	5.4	45.3	87.6

¹⁶ Las formas enumeradas corresponden al patrón de la expresión regular: $\zeta(\text{hm}|\text{hum}|\text{mhm}|\text{uhum}|\text{m+})\text{¥}?$.

4. Presentación informática

4.1. PRESEEA en LYNEAL

En la actualidad nadie duda de la utilidad de la estadística y de la informática en los estudios lingüísticos. Afortunadamente en los últimos años, gracias a los datos preparados en los grandes proyectos de investigaciones de lingüística de variación, estamos cada vez más capacitados para llevar a cabo estudios de gran envergadura. El proyecto PRESEEA ofrece unos datos de primera calidad en el sentido de que están bien planeados: cuentan con una estructura idónea para analizar relaciones entre las formas lingüísticas y los atributos extralingüísticos.

La calidad de los estudios lingüísticos se ha mejorado no solamente en los materiales de análisis, sino también en los tratamientos estadísticos e informáticos. Para aplicar las teorías probabilísticas al tratamiento informático, hemos desarrollado un interfaz relativamente fácil de manejar: el sistema LYNEAL. Está funcionando en los dos sitios, uno en Madrid y otro en Tokio (véase 4.1):

<http://shimoda.llf.uam.es/ueda/lyneal/preseea.htm>

<http://lecture.ecc.u-tokyo.ac.jp/~cueda/lyneal/preseea.htm>

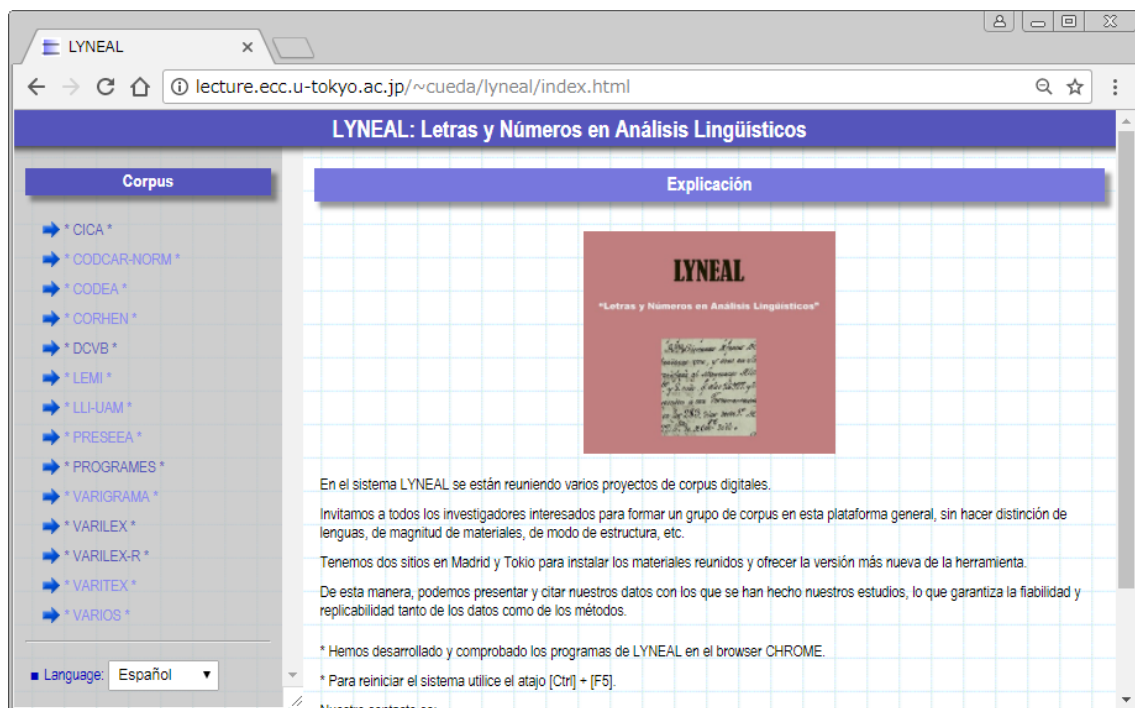


Figura 4.1.a Sistema LYNEAL

Los rasgos característicos del sistema LYNEAL son:

- facilitar el modo fácil de buscar las formas utilizando expresiones regulares simplificadas
- posibilitar el reemplazo de textos a la hora de búsqueda, de modo que las etiquetas innecesarias pueden ser eliminadas
- filtrar los atributos necesarios, por ejemplo, lugar, papel, sexo, edad, nivel de educación
- combinar los múltiples atributos para analizar de manera separada y/o unida
- seleccionar los componentes deseados
- elaborar distintos tipos de frecuencia (absoluta, relativa, normalizada, probabilística, tipificada, etc.)
- calcular distintos valores estadísticos básicos (media, varianza, desviación típica, mediana, cuartiles, rango, media proporcional, etc.)
- posibilitar análisis multivariantes
- ofrecer distintos tipos de gráficos

En lo siguiente explicaremos las funciones básicas que se encuentran en el interfaz de PRESEEA en LYNEAL.

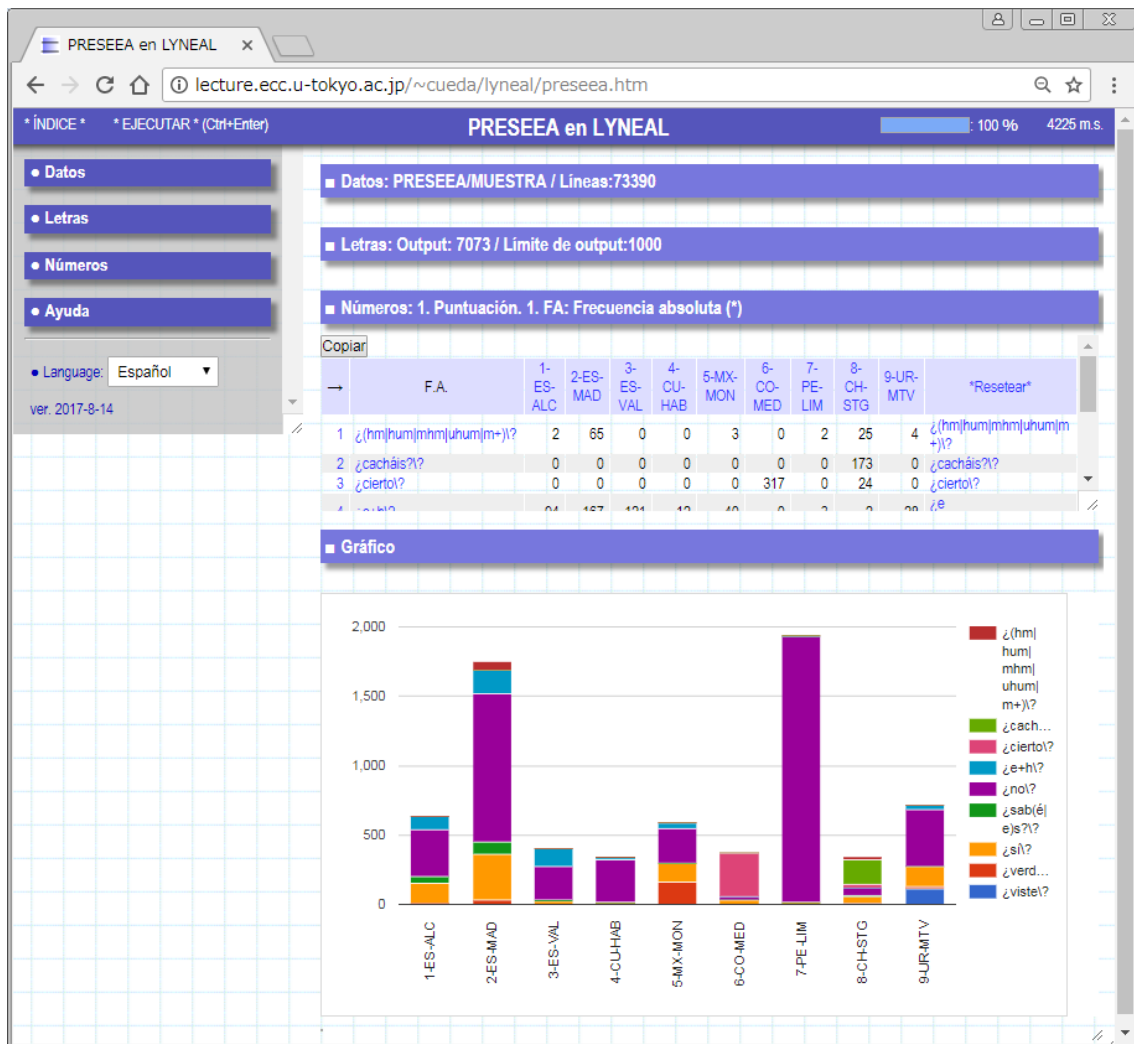
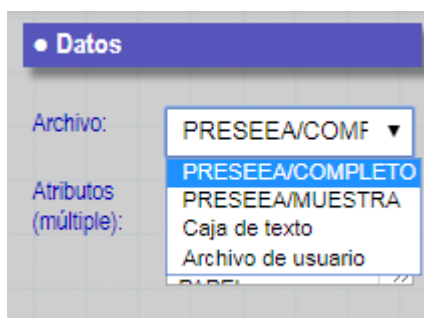


Figura 4.1.b PRESEEA en LYNEAL

<http://lecture.ecc.u-tokyo.ac.jp/~cueda/lyneal/preseea.htm>

4.2. Datos



En la zona izquierda de la página web, se encuentra en primer lugar la sección de [Datos], donde el usuario puede escoger una de 4 opciones:

- PRESEEA/COMPLETO: Textos completo del Corpus PRESEEA
- PRESEEA/MUESTRA: Textos de muestras del Corpus PRESEEA
- Caja de texto: donde el usuario puede pegar los textos anteriormente copiados
- Archivo de usuario: donde el usuario puede subir el archivo de texto (*.txt) que se encuentra en su entorno local (ordenador).

4.3. Letras

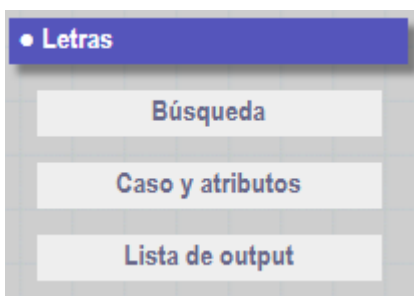


Fig. 4.3.1. Letras

En la sección de [Letras] se encuentran tres subsecciones: [Búsqueda], [Caso y atributos] y [Lista de output]. Dando un clic a [Búsqueda], se despliega:

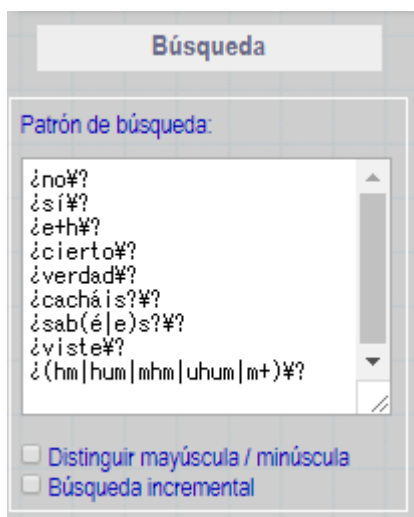


Fig. 4.3.2. Letras - Búsqueda

donde escribimos los patrones de búsqueda en el lenguaje de Expresión Regular¹⁷:

¿no%?
 ¿sí%?
 ¿e+h%?

¹⁷ El lenguaje de búsqueda en la Explicación Regular está explicado en la [Ayuda].

- ¿cierto?¥?
- ¿verdad?¥?
- ¿cachái(s)?¥?
- ¿sab(é|e)s?¥?
- ¿viste?¥?
- ¿(hm|huh|mhm|uhum|m+)¥?

Seguidamente, abrimos la subsección de [Caso y atributos]:

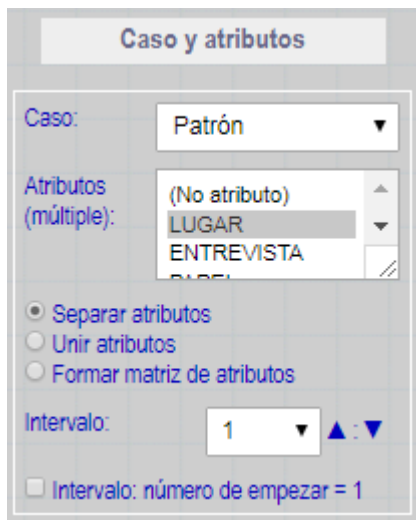


Fig. 4.3.3. Letras - Casos y atributos

donde seleccionamos en la cajilla de [Atributos (múltiples)] LUGAR.

Ejecutamos pulsando [EJECUTAR], que se encuentra en la parte superior izquierda o tecleando el atajo [Ctrl] + [Enter] y sale la tabla siguiente en la parte de [Números] en la zona izquierda:

→	F.A.	1-ES-ALC	2-ES-MAD	3-ES-VAL	4-CU-HAB	5-MX-MON	6-CO-MED	7-PE-LIM	8-CH-STG	9-UR-MTV
1	¿(hm huh mhm uhum m+)?	2	65	0	0	3	0	2	25	4
2	¿cacháis?!	0	0	0	0	0	0	0	173	0
3	¿cierto?!	0	0	0	0	0	317	0	24	0
4	¿e+hl?	94	167	121	12	40	0	3	2	28
5	¿no!?	340	1068	243	308	248	21	1914	59	407
6	¿sab(é e)s?!	50	92	10	1	1	0	1	1	2
7	¿sí!?	139	329	21	5	135	33	13	52	147
8	¿verdad!?	12	30	3	8	162	0	1	1	18
9	¿viste!?	0	0	0	0	0	0	0	7	109

Fig. 4.3.4. Tabla de Frecuencia Absoluta-

Podemos copiar estos datos pulsando el botón de [COPIAR] que se encuentra

arriba de la tabla. Los datos copiados se pueden pegar en Excel o en Word.

4.4. Números

Debajo de la sección de [Letras], se encuentra la sección de [Números]:



Fig. 4.4.1. Números

Al dar clic a [Análisis numéricos] desplegamos la casilla siguiente:

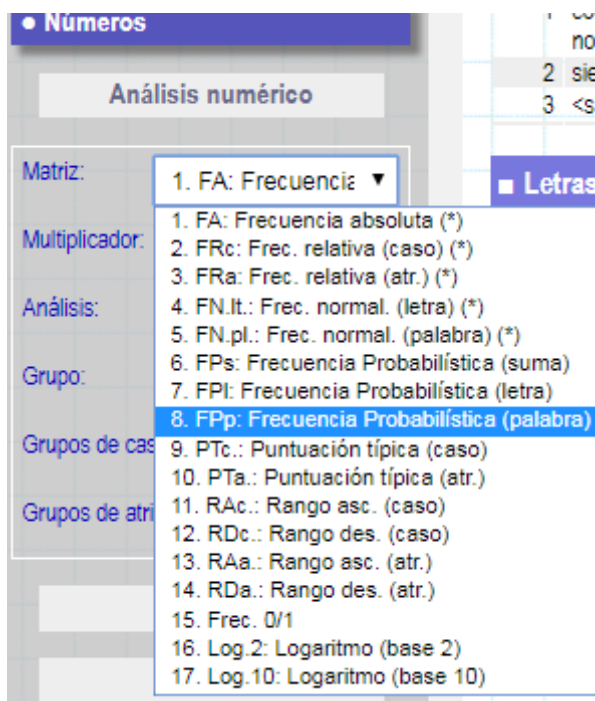


Fig. 4.4.2. Números - Matriz

donde seleccionamos 8: FpP: Frecuencia Probabilística (palabra). Y al mismo tiempo seleccionamos 100000 en el [Multiplicador]:



Fig. 4.4.3. Números - Multiplicador

De esta manera obtenemos la tabla de la Frecuencia Probabilística (palabra) multiplicada por 100 000:

→	FP.PI.:100000	1-ES-ALC	2-ES-MAD	3-ES-VAL	4-CU-HAB	5-MX-MON	6-CO-MED	7-PE-LIM	8-CH-STG	9-UR-MTV
1	¿(hm hum mhm uhum m+)?	.0	21.6	.0	.0	.0	.0	.0	4.4	.1
2	¿cacháis?!	.0	.0	.0	.0	.0	.0	.0	64.9	.0
3	¿cierto?!	.0	.0	.0	.0	.0	189.9	.0	4.1	.0
4	¿e+hl?	42.6	62.9	64.4	1.6	7.1	.0	.0	.0	7.2
5	¿no?!	177.2	464.9	139.6	165.2	71.8	5.0	1047.7	19.1	228.0
6	¿sab(é e)s?!	19.8	32.2	1.2	.0	.0	.0	.0	.0	.0
7	¿sí?!	66.2	132.3	4.9	.2	36.6	10.2	1.7	16.2	74.7
8	¿verdad?!	1.5	5.9	.0	.7	44.9	.0	.0	.0	3.5
9	¿vistel?!	.0	.0	.0	.0	.0	.0	.0	.3	53.4

Fig. 4.4.4. Tabla de Frecuencia probabilística multiplicada por 100 000 palabras

5. Final

En la frecuencia de las distintas formas de la función discursiva de pregunta confirmativa, lo mismo que suele ocurrir en las frecuencias de los fenómenos lingüísticos en general (Juilland and Chang-Rodriguez, 1964; Baayen, 2001; Davies, 2006; Taiylor, 2012), se averigua una distribución extremadamente sesgada: unas pocas formas, *¿no?*, *¿sí?*, *¿eh?*, *¿cierto?*, *¿verdad?*, etc., se repiten con la frecuencia sumamente alta, mientras que la mayoría son de frecuencia bastante reducida. Para este tipo de distribución, no sirve la Media ni la Mediana como valor representativo del conjunto. En su lugar hemos propuesto el uso de la nueva fórmula de Media

Proporcional.

Por otra parte, sabemos que la Frecuencia Absoluta (neta) no sirve para un estudio comparativo, puesto que la base suele ser diferente, es decir, el número de palabras en cada hablante o parámetro es diferente al de los otros. De ahí que recurramos a la Frecuencia Relativa o a la Frecuencia Normalizada por cierta cantidad de palabras. A la hora de efectuar tal relativización o normalización tampoco estamos seguros de su utilidad, puesto que la base suele ser muy diferente según los parámetros seleccionados. Por esta razón, hemos intentado aplicar la teoría de la probabilidad binomial y hemos encontrado la solución de esta falta de comparabilidad en forma de Frecuencia Probabilística.

Finalmente, hemos propuesto un método no muy usual de Microanálisis de frecuencias dotadas de múltiples variables, al lado del método habitual de Macroanálisis. La razón por la que recomendamos el método de Microanálisis está en que ahí se puede observar unos hechos importantes de distribución, difíciles de encontrar con el método de Macroanálisis.

Tanto el cálculo de la Media Proporcional y de la Frecuencia Probabilística como el método de Microanálisis son difíciles de realizar, por no decir imposibles, sin recurrir a la herramienta informática. Hemos buscado la combinación ideal entre la parte lingüística con la estadística en la plataforma informática común, que funciona de manera lineal desde la selección de los datos lingüísticos, pasando por la búsqueda general, para llegar a las tablas de distribución y la visualización gráfica. De esta manera ofrecemos la misma herramienta al servicio de los investigadores interesados en los cálculos estadísticos con los que lleven a cabo los estudios con evidencias replicables.

Referencias

- Asociación de Academias de la Lengua Española. (2010) *Diccionario de americanismos*. Madrid: Santillana Ediciones.
- Baayen, R. H. (2001) *Word frequency distributions*. Dordrecht. Kluwer Academic Publishers.
- Baixauli Fortea, I. (2000) "Las secuencias de historia" en Briz, A. / Grupo Val.Es.Co. *¿Cómo se comenta un texto coloquial?* Barcelona. Ariel. p. 81-107.
- Beinhauer, W. (1978) *El español coloquial*. Madrid. Gredos.
- Bolinger, D.L. / Bowen J. D. / Brady, A. M. / Haden, E. F. / Poston, L./ Sacks N. P. (1973) *Modern Spanish. Third Edition. A project of Modern Language Association* .

- New York. Harcourt Brace Jovanovich, Inc.
- Briz Gómez, A. (1998) *El español coloquial en la conversación. Esbozo de pragmática*. Barcelona: Ariel.
- _____. / Grupo Val.Es.Co. Universidad de Valencia (2000) *Diccionario de partículas discursivas del español* <http://www.dpde.es/>
- Casado Velarde, M. (1998) "Lingüística del texto y marcadores del discurso", en Martín Zorraquino, M. A. / Montolío Durán, E. *Los marcadores del discurso. Teoría y análisis*. Madrid Arco / Libros.
- Cestero Mancera, A. M. (2000) "La función fática del lenguaje en el discurso y en la conversación", *Actas del IV Congreso de Lingüística General*, Cádiz, vol. II, p. 617-629.
- _____. (2003) "El funcionamiento de los apéndices interrogativos en la conversación y en el discurso académico", Castillo Martínez, Cristina / Lucía Megías, J. M. (eds.) *Decíamos ayer... Estudios de alumnos en honor a María Cruz García de Enterría*. Universidad de Alcalá. p. 83-127.
- Cortes Rodríguez, L. / Camacho Adarve, M. M. (2005) *Unidades de segmentación y marcadores del discurso: Elementos esenciales en el procesamiento discursivo oral*. Madrid. Arco / Libros.
- Fuentes Rodríguez, C. (1990) "Apéndices con valor apelativo", P. Carbonero Cano (coord.) y M.T.Palet Plaja (ed.) *Sociolingüística andaluza, 5, Habla de Sevilla y hablas americanas*. Sevilla: Universidad de Sevilla, pp.171-196.
- Gelabert, M. J. / Herrera, M. / Martinell, E. / Martinell, F. (1993) *Repertorio de funciones comunicativas del español*. Madrid. SGEL. Traducción japonesa por Hara M. / Eto, I. Tokio. Sanshuusha.
- Haensch, G. / Werner, R. (1993) *Nuevo diccionario de americanismos*. Tomo I. *Nuevo diccionario de colombianismos*. Bogotá. Instituto Caro y Cuervo.
- _____. / Köhl de Mones, U. (1993) *Nuevo diccionario de americanismos*. Tomo III. *Nuevo diccionario de uruguayismos*. Bogotá. Instituto Caro y Cuervo.
- Juillard, A. / Chang-Rodríguez, E. (1964) *Frequency dictionary of Spanish Words*. The Hague. Mouton.
- Davies, Mark. 2006. *A frequency Dictionary of Spanish. Core vocabulary for learners*. London. Routledge.
- leanManufacturing. *ABC Analysis* (acceso: 14/8/2017):
<http://leanmanufacturingpdf.com/lean-logistics-abc-analysis/>
- Lawaetz, G. (1972) *Spanish short stories. vol. 2. Cuentos hispánicos, t.2*. Penguin Books. Middlesex (England).

- Moreno Fernández, F. (2005) «Corpus para el estudio del español en su variación geográfica y social. El corpus PRESEEA», *Oralia*, 8. p. 123-139.
- Steel, B. (1985) *A textbook of colloquial Spanish*. Madrid. Sociedad General Española de Librería.
- Quilis, A. (1981) *Fonética acústica de la lengua española*. Madrid. Gredos. p.440-442.
- ____ (1993) *Tratado de fonología y fonética españolas*. Madrid. Gredos. p. 451-452.
- Taiylor, J. R. (2012) *The Mental Corpus. How Language is Represented in the Mind*. Oxford University Press.
- Takahashi, K. (1998) *Diccionario temático de expresiones en español*. Tokio. Hakusuisha. (en japonés).
- Ueda, H. (2017) *Análisis de datos cuantitativos para estudios lingüísticos*:
<http://lecture.ecc.u-tokyo.ac.jp/~cuedágengó4-numeros/doc/numeros-es.pdf>

ADDENDA-1: Media proporcional

Problema

Para representar el conjunto de cifras, solemos utilizar el promedio, que se obtiene por la división de la totalidad por el número de datos. Por ejemplo, el promedio del conjunto $\{1, 3, 5, 8, 10\}$ es $(1 + 3 + 5 + 8 + 10) / 5 = 5.4$. En Estadística general este promedio se denomina «Media Aritmética» (MA). La Media Aritmética (MA) es apropiada para ser aplicada al conjunto de datos sin valores extraordinarios. En cambio, cuando se aplica al conjunto que los contiene, por ejemplo $\{1, 1, 2, 3, \underline{153}\}$ donde apreciamos un valor extraordinario, 153, la Media Aritmética (MA), $(1 + 1 + 2 + 3 + 153) / 5 = 32$, presenta un valor muy distante de la realidad del conjunto. Es decir, este valor (32) dista tanto de $\{1, 1, 2, 3\}$ como de 153, de modo que no podemos confirmar su representatividad. En tal caso, se recomienda en la estadística general el uso de Mediana, que es el valor que se encuentra en el punto medio en la serie ordenada, $(1, 1, \underline{2}, 3, 153)$. Efectivamente, en el punto medio se encuentra 2, que es la Mediana de este conjunto. Sin embargo, esta Mediana también posee un defecto, es decir, el valor de la mediana 2 no significa mucho o casi nada para los miembros altos o bajos distantes, 153. Por consiguiente, la Mediana no sirve tampoco para representar el conjunto de datos.

Solución

Por lo tanto, para un conjunto de datos que contienen unos valores extraordinarios, proponemos utilizar una clase de la media que denominamos «Media

Proporcional» (MP) (Ueda 2017). Para explicarla, primero repasamos la fórmula de la Media Aritmética (MA) y, seguidamente, explicamos la fórmula de la Media Proporcional (MP) de manera comparativa.

$$\begin{aligned} \text{MA} &= (x_1 + x_2, \dots, x_n) / n && n: \text{recuento de casos} \\ &= 1/n (x_1 + x_2, \dots, x_n) \\ &= 1/n \sum x_i \end{aligned}$$

La fracción $1/n = 1/5$ se la puede distribuir de la manera siguiente:

$$\begin{aligned} \text{MA} &= 1/n (x_1 + x_2, \dots, x_n) \\ &= (x_1/n + x_2/n, \dots, x_n/n) \end{aligned}$$

Concretamente:

$$= (1/5 + 1/5 + 2/5 + 3/5 + 153/5) = 32$$

De esta manera, la Media Aritmética (MA) equivale al producto de suma de las multiplicaciones de la frecuencia por el peso constante $1/n (= 1/5)$.

En cambio, en el cálculo de la Media Proporcional (MP), en lugar del valor constante de $1/n (= 1/5)$, utilizamos la ratio de la frecuencia con respecto a la totalidad de la frecuencia (T) de manera siguiente:

$$\begin{aligned} \text{T} &= (x_1 + x_2, \dots, x_n) = 1 + 1 + 2 + 3 + 153 = 160 \\ \text{MP} &= (x_1 x_1 / T + x_2 x_2 / T, \dots, x_n x_n / T) \end{aligned}$$

Concretamente:

$$= 1*1/160 + 1*1/160 + 2*2/160 + 3*3/160 + 153*153/160$$

De esta manera, estamos ponderando cada cifra multiplicándola por la proporción que ocupa dentro de la totalidad ($T = 160$). Este cálculo se simplifica de la manera siguiente:

$$= (1^2 + 1^2 + 2^2 + 3^2 + 153^2) / 160$$

Es decir:

$$\begin{aligned} &= (x_1 x_1 + x_2 x_2, \dots, x_n x_n) / T \\ &= (x_1^2 + x_2^2, \dots, x_n^2) / T \end{aligned}$$

Concretamente:

$$= (1^2 + 1^2 + 2^2 + 3^3 + 153^2) / 160$$

$$= 146.4$$

Notamos que el valor de la Media Proporcional (MP), 146.4, refleja la magnitud del conjunto de datos {1, 1, 2, 3, 153}. Compárense la Suma, la Máxima, la Media Aritmética (MA) y la Media Proporcional (MP) en la siguiente tabla (X):

X	v1	v2	v3	v4	v5	Suma	Máxima	MA	MP
d1	10	19	14	7	12	62	19	12.4	13.7
d2	11	7	10	0	1	29	11	5.8	9.3
d3	0	0	1	12	1	14	12	2.8	10.4
d4	0	1	2	3	3	9	3	1.8	2.6

Como la Media Proporcional (MP) representa la magnitud del conjunto de datos, se acerca a la Máxima alejándose de los valores mínimos. Esta representatividad de la magnitud constituye la característica principal de la Media Proporcional (MP), a diferencia de la Media Aritmética (MA), que suele presentar la cifra igualitaria, pero distante tanto de las frecuencias menores que están en la cola, como de las mayores en la cabeza del conjunto caracterizado por la distribución de tipo de la letra L y, por otra parte, a diferencia de la Mediana, que ofrece la cifra que represente la mayoría de los muchos miembros menores del mismo conjunto del tipo de la letra L, que no tiene mucho que ver con la magnitud misma del conjunto.

La siguiente tabla muestra que la Media Proporcional (MP) representa correctamente la magnitud del conjunto vertical de cada columna, que se aproxima a la Máxima, mientras que la Media Aritmética (MA) falla en representar tanto los valores mayores como los menores del conjunto de las frecuencias que tienen valores sumamente distantes entre sí y, por la Mediana está demasiado a favor de los menores:

.F.N.PL: 100.000	Es.A.	Es.M.	Es.V.	Cu.H.	Mx.M.	Co.M.	Pe.L.	Ch.S.	Ur.M.
1. ¿no?	267.6	660.3	223.9	231.8	116.7	16.9	1502.6	34.1	323.9
2. ¿sí?	111.1	217.8	21.1	3.8	65.2	26.6	11.7	31.8	124.7
3. ¿eh?	73.6	102.8	112.9	9.0	19.2	0	2.3	1.2	22.0
4. ¿cierto?	0	0	0	0	0	255.1	0	13.9	0
5. ¿verdad?	9.4	18.5	2.8	6	76	0	0.8	0.6	14.1
6. ¿cachái(s)?	0	0	0	0	0	0	0	100.6	0
7. ¿sabes?	39.1	56.6	9.2	0.8	0.5	0	0.8	0.6	1.6
8. ¿viste?	0	0	0	0	0	0	0	4	86.3
9. ¿hm?	1.6	40	0	0	1.4	0	1.6	14.5	3.1
Suma	502.4	1096	369.9	251.4	279.0	298.6	1519.8	201.3	575.7
Máxima	267.6	660.3	223.9	231.8	116.7	255.1	1502.6	100.6	323.9
MA	55.8	121.8	41.1	27.9	31.0	33.2	168.9	22.4	64.0
Mediana	9.4	40.0	2.8	.8	1.4	.0	.8	13.9	14.1
MP	181.1	455.4	171.4	214.3	86.1	221.3	1485.7	63.2	223.4

Como hemos visto, la Media Proporcional (MP) se aproxima a la Máxima, de modo que la Máxima, fácil de calcular, también podría servir aparentemente como representante de la magnitud del conjunto, en lugar de la Media Proporcional (MP). No obstante, por ejemplo, la Máxima del conjunto $\{1, 1, 2, 3, 153\}$ y la de $\{10, 10, 20, 30, 153\}$ resultan ser la misma (153), haciendo caso omiso de los valores menores $\{1, 1, 2, 3\}$, $\{10, 10, 20, 30\}$, mientras que la Media Proporcional (MP) de los dos conjuntos son 146.4 y 111.7, respectivamente, es decir, los valores menores se toman en consideración de manera proporcional. La mayor cifra de Media Proporcional (MP) del primer conjunto, $MP(1, 1, 2, 3, 153) = 146.4$, con respecto al segundo, $MP(10, 10, 20, 30, 153) = 111.7$, es el reflejo de la mayor concentración o destacamiento de la Máxima (153) del primer conjunto, mientras que en el segundo se ha aproximado a los valores menores que tienen cierta fuerza. La característica media de la Media Proporcional (MP) se comprueba en los tres conjuntos $\{0, 0, 0, 0, 153\}$, $\{0, 0, 0, 153, 153\}$, $\{0, 0, 153, 153, 153\}$, que dan todos la cifra MP de 153, lo mismo que $\{153\}$, $\{153, 153\}$, $\{153, 153, 153\}$, que dan igualmente la cifra MA de 153.

ADDENDA-2: Frecuencia probabilística

Problema

En la lingüística de corpus, con la búsqueda múltiple (de varias formas al mismo tiempo) con atributos también múltiples (por ejemplo, años), llegamos a obtener una tabla bidimensional, de formas y de años. Ahora estamos en condiciones de ver el fenómeno no solamente en una franja de años (1200, siglo XIII), sino de observar los cambios lingüísticos a lo largo de la cronología (1200, 1250, etc.) en comparación con otras formas. Veamos un dato real de la Frecuencia Absoluta (FA) de las tres formas con variación ortográfica: <uoz>, <voz>, <boz>¹⁸:

FA	1200	1250	1300	1350	1400
<i>uoz</i>	3	8	3	11	6
<i>boz</i>	0	3	8	18	35
<i>voz</i>	0	1	1	23	53
Sm	3	12	12	52	94

Estas frecuencias, sin embargo, no son comparables, puesto que las Sumas (Sm) de las tres formas son diferentes {3, 12, 12, 52, 94}. Por ejemplo, la cifra 3 <uoz> en 1200 no es comparable con la 8 de la misma forma en 1250. En tal caso, los investigadores recurren a la Frecuencia Relativa (FR), que se calcula por la división de la Frecuencia Absoluta por la Suma, por ejemplo, $3 / 3 = 1.000$, $8 / 12 = .667$. Si multiplicamos la Frecuencia Relativa por 100, llegamos a la cifra de porcentaje: $1.000 * 100 = 100$ (%), $0.667 * 100 = 66.7$ (%):

FA	1200	1250	1300	1350	1400	FR (%)	1200	1250	1300	1350	1400
<i>uoz</i>	3	8	3	11	6	<i>uoz</i>	100.0	66.7	25.0	21.2	6.4
<i>boz</i>	0	3	8	18	35	<i>boz</i>	0.0	25.0	66.7	34.6	37.2
<i>voz</i>	0	1	1	23	53	<i>voz</i>	0.0	8.3	8.3	44.2	56.4
Sm	3	12	12	52	94	Sm	100.0	100.0	100.0	100.0	100.0

Sin embargo, ni la Frecuencia Relativa (FR) ni el porcentaje (%) son adecuados para comparar las cifras con bases distintas. Por ejemplo, 3 entre 3 (FR: 1.000 (100%)) presenta la cifra mayor que 8 entre 12 (FR: 0.667 (66.7%)), a pesar de que pensamos e intuimos que 3 entre 3 es menos importante que 8 entre 12 y mucho menos importante

¹⁸ La tabla se ha obtenido en el sitio de «CODEA en LYNEAL»: <http://shimoda.llif.uam.es/ueda/lyneal/codea.htm>

que 80 entre 120. Los aficionados de fútbol saben o intuyen que el futbolista que ha metido 3 goles en 3 partidos es menos importante que el otro ha metido 8 goles entre 12. Todo esto significa que el porcentaje no sirve para la comparación numérica, por la razón de que, por ejemplo, 3 goles en 10 partidos no garantiza 30 goles en 100 partidos, lo que dice precisamente el 30%. Creemos que el porcentaje sirve para describir la proporción que ocupa cada caso dentro del conjunto, pero no sirve para comparar cada caso entre varios conjuntos. Más adelante buscaremos la solución de este problema de la evaluación numérica, propio de la Frecuencia Relativa (FR) y del porcentaje.

Pero antes, veamos el problema de otra frecuencia también utilizada en la lingüística de corpus en general. Se trata de la Frecuencia Normalizada (FN), que se calcula por la división de la Frecuencia Absoluta (FA) por la Totalidad de Palabras (TP) contadas en cada sección, multiplicada por algún Multiplicador (M) apropiado.

$$FN = FA / TP * M$$

Por ejemplo, en la franja de 1200 del corpus se han contado 7 736 palabras en total, que es la Totalidad de Palabras (TP). Entonces, la Frecuencia Normalizada de <uoz> en 1200 es $3 / 7\ 736 * 100\ 000 = 38.8$. Recomendamos utilizar como Multiplicador (M) el número redondeado próximo a la Máxima de la base (TP): 96 059 (en el conjunto de datos de 1400). Llegamos a obtener la tabla inferior derecha (FN):

FA	1200	1250	1300	1350	1400
uoz	3	8	3	11	6
boz	0	3	8	18	35
voz	0	1	1	23	53
TP	7 736	36 052	40 957	64 999	96 059

FN.	1200	1250	1300	1350	1400
uoz	38.8	22.2	7.3	16.9	6.2
boz	0.0	8.3	19.5	27.7	36.4
voz	0.0	2.8	2.4	35.4	55.2

Sin embargo, aquí en la Frecuencia Normalizada (FN) también se presenta el mismo problema de falta de comparabilidad propia de los datos de bases diferentes, y especialmente, de algunas bases bastante reducidas. No podemos menos de sentir dudas sobre la cifra FN de <uoz> en 1200, 3 entre 7 736 cuya FN es 38.8 en comparación con la FN de la misma forma <uoz> en 1250, 8 entre 36 052 cuya FN es 22.2. Nos preguntamos si 38.8 es realmente comparable con 22.2.

La esencia del problema es la misma tanto en la Frecuencia Relativa (FR) como en la Frecuencia Normalizada (FN) en el sentido de que las dos calculan sobre bases diferentes. Paradójicamente, las dos frecuencias se utilizan precisamente cuando las bases son diferentes, puesto que si las bases son iguales no hace falta recurrir a estas

frecuencias y en la Frecuencia Absoluta neta podemos hacer la comparación numérica sin problema.

El problema de la falta de comparabilidad tratado en este apartado se soluciona por la eliminación del conjunto en cuestión. En el ejemplo de los datos de las tres formas medievales, se trataría de eliminar el conjunto correspondiente a 1200. Es la práctica general en el tratamiento estadístico. Por ejemplo, en el mundo deportivo de béisbol, se calcula la puntuación de los jugadores con participación suficiente en los partidos. Los jugadores que no pasan el umbral establecido están excluidos de la evaluación desde el principio. Pero nos preguntamos qué hacemos con la franja de 1250, donde se registran las frecuencias dentro de la base de casi un tercio de 1400 (37.5%). Nuestra idea es tratar todos los datos sin distinción, pero con criterios comunes de probabilidad. Nuestro método, que a continuación vamos a explicar, ofrece la evaluación de los datos de manera equitativa, con bases similares o distantes, lo que demuestra la robustez absoluta, en comparación con los métodos tradicionales de la Frecuencia Relativa (FR), inclusive su variante Porcentaje, y la Frecuencia Normalizada (FN), cuya fragilidad hemos visto en los casos de bases distantes en esta sección.

Solución

Para ofrecer la solución del problema propio de la Frecuencia Relativa (FR) y de la Frecuencia Normalizada (FN), presentamos una nueva fórmula de frecuencia. Nuestro propósito es buscar un tipo de frecuencia, «Frecuencia Probabilística» (FP), que represente el valor relativo de la Frecuencia Absoluta (FA) dentro del conjunto (base) con unos cálculos sencillos de la probabilidad (Ueda 2017). Se justifica por los resultados de experimentos reales y por nuestra intuición y pensamiento que, por ejemplo, 30 entre 100 es más importante (o significativo) que 3 entre 10, aun siendo ambos iguales de 0.3 (30%) en la probabilidad de ocurrencia. Para demostrarlo, recurrimos a la fórmula probabilística binomial. El camino para llegar a conocer la Frecuencia Probabilística (FP) recorre los tres pasos siguientes: (1) Significatividad (S), (2) Probabilidad Esperada (PE) y, finalmente, (3) Multiplicador (M).

(1) Significatividad

Intuimos y pensamos que, por ejemplo, un futbolista que ha metido 28 goles en 100 partidos es más «importante» y ha contribuido más al equipo que el otro que ha metido 3 goles en 10 partidos, aunque la ratio de goles del primero ($28 / 100 = 28\%$) es menor que la del segundo ($3 / 10 = 30\%$). Al grado de importancia, damos el nombre de

«Significatividad» (S) (ing. *significance*). Partimos de unos casos sencillos y especiales para llegar al caso general, aplicable a las frecuencias en general.

Para calcular la Significatividad (S), utilizamos la probabilidad. Para llegar a entenderla, empezamos con unos ejemplos tan sencillos como de tirada de una moneda, donde cada faz tiene la Probabilidad Esperada (PE) de 0.5 (50%)¹⁹.

La tabla siguiente muestra las probabilidades de dos eventos ($x = 0, 1$) de una tirada (número de ensayo $n=1$) de una moneda: «cara», con valor de 1, o «cruz», con valor 0. La Probabilidad Esperada (PE) de cada una es 0.5, puesto que hay dos posibilidades: cara (1) o cruz (0). En la tabla cada evento viene con su propia Probabilidad de Ocurrencia (PO), que acabamos de ver, la Probabilidad Cumulativa (PC), que se va acumulando con cada Probabilidad de Ocurrencia correspondiente, y la Significatividad (S):

x	Caso	$PO(x, 1, 0.5)$	$PC(x, 1, 0.5)$	$S(x, 1, 0.5)$
cruz: $x = 0$	(0)	$1/2 = 0.5$	0.5	0
cara: $x = 1$	(1)	$1/2 = 0.5$	$0.5 + 0.5 = 1.0$	0.5

La columna de Probabilidad de Ocurrencia $PO(x)$ muestra en la primera fila la probabilidad de cruz ($x = 0$) que es $PO(0) = 1/2 = 0.5$ y, en la segunda fila, la de cara ($x = 1$) que es $PO(1) = 1/2 = 0.5$. La Probabilidad Cumulativa (PC) de $x = 0$, $PC(0)$, es 0.5, que es igual a $PO(0)$ y la de $x = 1$, $PC(1)$, es 1.0, que es suma de $PO(0) = 0.5$ y $PO(1) = 0.5$. La última Probabilidad Cumulativa (PC) es siempre 1.

Ahora bien, definimos la «Significatividad» (S) de x , $S(x, n, e)$, como correspondiente a la Probabilidad Cumulativa de $x - 1$, $PC(x-1, n, e)$:

$$S(x, n, e) = PC(x-1, n, e)$$

(x : Ocurrencia; n : Ensayos; e : Probabilidad Esperada)

La Significatividad (s) de $x = 0$, $S(0)$, la definimos como 0, por no existir la Probabilidad Cumulativa de -1:

$$S(0) = 0$$

La razón por la que consideramos la Significatividad como la Probabilidad

¹⁹ Esto significa que si tiramos una moneda, sale siempre una de las dos caras y si tiramos 1000 veces, casi la mitad de las veces (aproximadamente 500 veces), sale la cara y otra mitad de las veces, la cruz. Entonces intuimos y pensamos lógicamente que la Probabilidad Esperada (PE) de cara es 0.5 (50%).

Cumulativa de la Ocurrencia del caso inmediatamente anterior está en que pensamos que la suma de las probabilidades anteriores de la Ocurrencia correspondiente es la probabilidad de la significatividad de los números de ocurrencias anteriores. Si echamos una moneda, la Significatividad de la ocurrencia de 1 («cara») es 0.5, lo que es complementaria del Riesgo (no «cara», es decir, «cruz»), que es también 0.5. Por consiguiente,

$$\text{Significatividad} + \text{Riesgo} = 1$$

lo que quiere decir que hay Significatividad de 0.5 (50%) de la aparición de la cara y hay Riesgo de 0.5 (50%). Esto quiere decir que si apostamos por la aparición de la cara, hay un 50% de riesgo, lo que sabemos e intuimos sin recurrir a la teoría de probabilidad.

Hasta aquí hemos visto un caso muy sencillo en que tiramos la moneda solo una vez. ¿Qué ocurre si tiramos la misma moneda dos veces? La tabla siguiente muestra la distribución de Probabilidades de Ocurrencia (PO) que presentan en dos ensayos de tirar una moneda ($n = 2$). Hay tres casos posibles: $x = 0, 1, 2$, es decir, $(0,0)$, $(1,0) + (0,1)$ y $(1,1)$:

x	Caso	PO($x, 2, 0.5$)	PC($x, 2, 0.5$)	S($x, 2, 0.5$)
$x = 0$	$(0, 0)$	$1/4 = 0.25$	0.25	0
$x = 1$	$(0, 1); (1, 0)$	$2/4 = 0.50$	$0.25 + 0.50 = 0.75$	0.25
$x = 2$	$(1, 1)$	$1/4 = 0.25$	$0.75 + 0.25 = 1.00$	0.75

Esta vez la Probabilidad Esperada (PE) de «cara» es igualmente 0.5. La columna de la Probabilidad de Ocurrencia (PO) muestra que la PO de 0 ocurrencias de cara, $PO(0)$, es 0.25 (cruz, cruz) = $(0, 0)$, es decir 1 de 4 casos. El total de casos son 4, porque se cuentan 4 casos siguientes: $(0, 0)$, $(0, 1)$, $(1, 0)$, $(1, 1)$. La probabilidad de 1 ocurrencia de cara, (cara, cruz) + (cruz, cara); $(1, 0) + (0, 1)$ es 0.5 (2 de 4 casos). Y, finalmente, la probabilidad de 2 ocurrencias de cara, (cara, cara); $(1, 1)$, $PO(2)$ es 0.25, que ocurre 1 de 4 casos. La columna de la Probabilidad Cumulativa (PC) presenta las probabilidades sumadas desde 0 hasta 2 en cada ocurrencia: $x = 0, 1, 2$.

La última columna de Significatividad (S) corresponde al caso anterior de la Probabilidad Cumulativa (PC). La última Significatividad (S) de $n = 2$ es $S(2) = 0.75$, que representa un aumento considerable con respecto al experimento anterior en el que se tiraba solo una vez la moneda: 0.5 ($n = 1$), lo que quiere decir que 2 entre 2 ($S = 0.75$) es mucho más «significativo» (importante) que 1 entre 1 ($S = 0.5$), aun siendo ambos iguales de 100% de Probabilidad Cumulativa (PC). Sin embargo, la Significatividad (S) todavía no llega más que a 0.75 (75%), lo que quiere decir que hay

0.25 (25%) de Riesgo.

Ahora bien, precisamos los 3 parámetros de la Significatividad (S): x : ocurrencias, n : total de las veces de ensayos, e : Probabilidad Esperada (PE) en forma de la función $S(x, n, e)$:

$$S(2, 2, 0.5) = PC(1, 2, 0.5) = 0.75$$

De la misma manera, la Significatividad (S) de $x = 1$ es:

$$S(1, 2, 0.5) = PC(0, 2, 0.5) = 0.25$$

Veamos el experimento de 3 tiradas de ensayo ($n = 3$):

x	Caso	PO($x, 3, 0.5$)	PC($x, 3, 0.5$)	S($x, 3, 0.5$)
$x = 0$	(0,0,0)	1/8 = .125	.125	0
$x = 1$	(1,0,0), (0,1,0), (0, 0, 1)	3/8 = .375	.500	.125
$x = 2$	(1,1,0), (1,0,1), (0,1,1)	3/8 = .375	.875	.500
$x = 3$	(1,1,1)	1/8 = .125	1.000	.875

La Significatividad (S) de la última ocurrencia ($x = 3$) ha aumentado en 0.875 y por consiguiente ahora el Riesgo ha disminuido en 0.125: $1 - 0.875 = 0.125$.

$$S(3, 3, 0.5) = PC(2, 3, 0.5) = 0.875 \text{ (87.5\%)}$$

Si apostamos que no salga 3 veces la cara, hay probabilidad de 87.5% de ganar la apuesta, que es la Significatividad (S); y el Riesgo de perder la apuesta es de 12.5%. Deberíamos aumentar la Significatividad (S) hasta, por lo menos, 95% (0.95) y, si es posible, hasta 99%, con los Riesgos de 5% o 1%, respectivamente. De esta manera perdemos la apuesta solo 1 de 20 veces, o 1 de 100 veces.

Veamos el experimento de 10 ensayos ($n = 10$):

x	PO($x, 10, 0.5$)	PC($x, 10, 0.5$)	S($x, 10, 0.5$)
$x = 0$.001	.001	.000
$x = 1$.010	.011	.001
$x = 2$.044	.055	.011
$x = 3$.117	.172	.055
$x = 4$.205	.377	.172
$x = 5$.246	.623	.377
$x = 6$.205	.828	.623

$x = 7$.117	.945	.828
$x = 8$.044	.989	.945
$x = 9$.010	.999	.989
$x = 10$.001	1.000	.999

Finalmente cuando $x = 9$, obtenemos la Significatividad $S(9, 10, 0.5) = 0.989$, superior a 95%, y la $S(10) = 0.999$, superior a 99%, lo que quiere decir que podemos presentar la cifra de 9 entre 10 con Significatividad (S) mayor de 95% y la de 10 entre 10 con Significatividad (S) mayor de 99%. En realidad, al echar la moneda 10 veces si salen 9 veces la cara de la moneda, la probabilidad total de las ocurrencias menores de 9 [0, 1, 2, ..., 8] se suma a 98.9%, que es bastante significativo. Es decir, con la Significatividad de 98.9% podemos afirmar que 9 entre 10 es significativo (importante). Es significativo o importante en el sentido de que 9 o 10 entre 10 ocurren solo con la probabilidad de $0.010 + 0.001 = 0.011$ (1.1%). De la misma manera, podemos afirmar que 10 entre 10 posee la Significatividad de 0.999 (99.9%). Compárense con los casos de 1 entre 1 (Significatividad de 50%), 2 entre 2 (75%) y 3 entre 3 (87.5%).

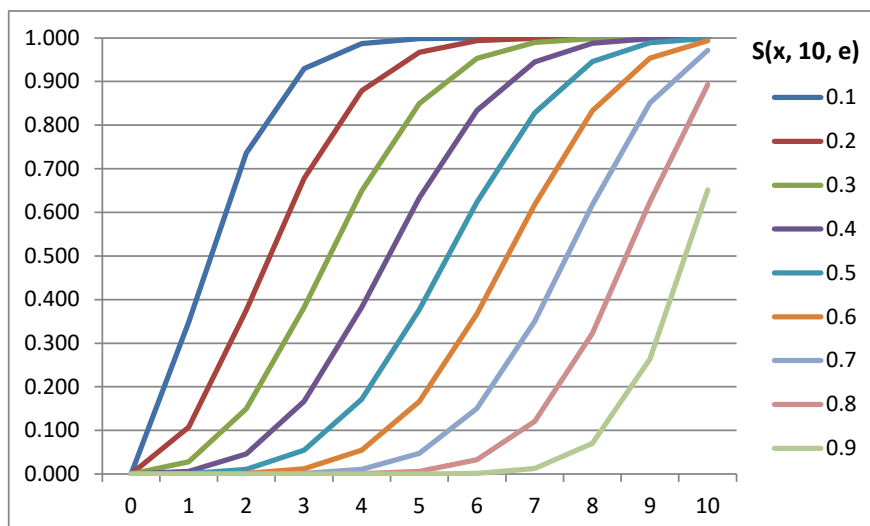
Hasta aquí hemos visto el comportamiento matemático de la Significatividad (S), que depende de los tres parámetros: x : ocurrencias, n : total de las veces de ensayos, e : Probabilidad Esperada (PE). Hemos observado su movimiento de acuerdo con x y n . Ahora veamos qué Significatividad (S) se presenta de acuerdo con el cambio de la Probabilidad Esperada (e). La tabla siguiente muestra la Significatividad (s) de las ocurrencias (x) de eventos dotados de la Probabilidad Esperada (e) de 0.1, por ejemplo, la Probabilidad Esperada (e) de sacar la tarjeta de «1» dentro de las diez tarjetas de {1, 2, ..., 10}:

x	PO($x, 10, 0.1$)	PC($x, 10, 0.1$)	S($x, 10, 0.1$)
$x = 0$.349	.349	.000
$x = 1$.387	.736	.349
$x = 2$.194	.930	.736
$x = 3$.057	.987	.930
$x = 4$.011	.998	.987
$x = 5$.001	1.000	.998
$x = 6$.000	1.000	1.000
$x = 7$.000	1.000	1.000
$x = 8$.000	1.000	1.000
$x = 9$.000	1.000	1.000
$x = 10$.000	1.000	1.000

Por ejemplo, cuando $x = 5$, $n = 10$, $e = 0.1$, $S(5, 10, 0.1)$ resulta ser 0.998, es decir, la suma de las Probabilidades de Ocurrencia (PO) $x = 0, 1, 2, 4$ es 0.998. Por consiguiente, al establecer la norma de la Significatividad (S) en 0.99, el 99% de las ocurrencias corresponden a 0, 1, 2, 3, 4. No aparece casi nunca 5 en adelante (5, 6, 7, ...) y existe una escasa probabilidad de 0.01 (1%).

La tabla siguiente ofrece las Significatividades (S) de acuerdo con las Ocurrencias ($x = 1, 2, \dots, 10$) y con las Probabilidades Esperadas ($e = 0.1, 0.2, \dots, 0.9$):

$S(x, 10, e)$	$e = 0.1$	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$x = 0$	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
1	0.349	0.107	0.028	0.006	0.001	0.000	0.000	0.000	0.000
2	0.736	0.376	0.149	0.046	0.011	0.002	0.000	0.000	0.000
3	0.930	0.678	0.383	0.167	0.055	0.012	0.002	0.000	0.000
4	0.987	0.879	0.650	0.382	0.172	0.055	0.011	0.001	0.000
5	0.998	0.967	0.850	0.633	0.377	0.166	0.047	0.006	0.000
6	1.000	0.994	0.953	0.834	0.623	0.367	0.150	0.033	0.002
7	1.000	0.999	0.989	0.945	0.828	0.618	0.350	0.121	0.013
8	1.000	1.000	0.998	0.988	0.945	0.833	0.617	0.322	0.070
9	1.000	1.000	1.000	0.998	0.989	0.954	0.851	0.624	0.264
10	1.000	1.000	1.000	1.000	0.999	0.994	0.972	0.893	0.651



Utilizamos la función de Excel BINOMDIST para obtener la Significatividad (S) de 0.349 en la celda de $x = 1$; $e = 0.1$:

$$s = S(x, n, e) = \text{BINOMDIST}(x-1, n, e, 1)$$

(s : Significatividad; x : Ocurrencia; n : Ensayos; e : Probabilidad Esperada)

$$0.349 \leftarrow S(1, 10, 0.1) = \text{BINOMDIST}(0, 10, 0.1, 1)$$

Esto quiere decir que al ensayar 10 veces del evento con la Probabilidad Esperada (PE) de 0.1, la ocurrencia 1 ($x = 1$) corresponde a la Significatividad (S) de .349. Las 2 ocurrencias ($x = 2$) del mismo evento corresponde a 0.736:

$$0.736 \leftarrow S(2,10,0.1) = \text{BINOMDIST}(1, 10, 0.1, 1)$$

(2) Probabilidad esperada

Hemos observado que la Significatividad (s) se obtiene por la función de $S(x, n, e)$ o la función de Excel BINOMDIST:

$$s = S(x, n, e) = \text{BINOMDIST}(x-1, n, e, 1)$$

(x : Ocurrencias, n : Ensayos, e : Probabilidad Esperada)

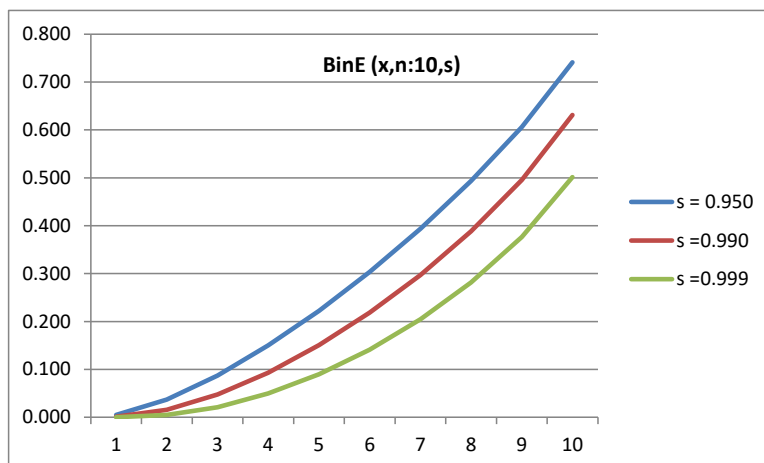
Por la función $S(x, n, e)$ se obtiene la Significatividad (s) por medio de x : Ocurrencias, n : Ensayos, e : Probabilidad Esperada . No obstante, en la práctica del análisis de datos lingüísticos, a diferencia de tales experimentos de la tirada de una moneda o el saque de una tarjeta, generalmente no se conoce la Probabilidad Esperada (PE) de los eventos desde el principio. Los parámetros que se conocen son x : Ocurrencias, n : Ensayos y la Significatividad (s) se establece por la parte del usuario. Por esta razón, seguidamente elaboramos la función E que devuelva la Probabilidad Esperada (PE) por medio de x :Ocurrencias, n :Ensayos y s : Significatividad:

$$e = E(x, n, s) = E(1, 10, 0.95)$$

La función $E(x, n, s)$ devuelve Probabilidad Esperada (PE) que se supone de un evento que ocurre x veces en n ensayos con la Significatividad s . Se trata de presuponer la Probabilidad Esperada (e) de, por ejemplo, 5 ocurrencias ($x = 5$) en 10 ensayos ($n = 10$) con la Significatividad (s) de, por ejemplo, 0.99 ($s = 0.99$). Con estos tres parámetros, la Probabilidad Esperada (PE) resulta calculable.

La tabla siguiente muestra las Probabilidades Esperada s (PE: e) de los eventos de 10 ensayos ($n = 10$), de acuerdo con las Ocurrencias (x) de 1 a 10 ($x = 1, 2, \dots, 10$), y con distintas Significatividades (s): $s = 0.95, 0.99, 0.999$. En esta tabla observamos que cuanto mayor es la Ocurrencia (x), tanto mayor es la Probabilidad Esperada (e). Por ejemplo, en la Significatividad (s) de 0.99, la Probabilidad Esperada (PE) de $x = 1$ es 0.001, mientras que la de $x = 10$ es 0.631.

$E(x, n:10, s)$	$s = 0.95$	$s = 0.99$	$s = 0.999$
$x = 1$	0.005	0.001	0.000
2	0.037	0.016	0.005
3	0.087	0.048	0.021
4	0.150	0.093	0.050
5	0.222	0.150	0.090
6	0.304	0.218	0.141
7	0.393	0.297	0.205
8	0.493	0.388	0.282
9	0.606	0.496	0.376
10	0.741	0.631	0.501



Al mismo tiempo confirmamos que el aumento de la Significatividad (s) causa la disminución de la Probabilidad Esperada (e). Por ejemplo la $E(5, 10, 0.95)$ es 0.222, mientras que la misma con la Significatividad (s) de 0.99 es 0.150 y la misma con la Significatividad de 0.999 es 0.090.

Supongamos que hemos tenido 2 veces éxito ($x = 2$) en 10 experimentos ($n = 10$). Con estos datos, sin embargo, no podemos esperar 20 éxitos en 100 experimentos futuros²⁰. Veamos cómo se presentan las Probabilidades Esperada (s) al aumentar el número de experimentos $n = 10, 100, 1000, \dots$:

n	$E(n*0.2, n, 0.99)$
$n = 10$	0.016
100	0.116

²⁰ Como veremos inmediatamente, incluso no podemos esperar 2 éxitos en 10 próximos experimentos.

1,000	0.171
10,000	0.191
100,000	0.197
1,000,000	0.199
10,000,000	0.200
100,000,000	0.200
1,000,000,000	0.200

En la tabla anterior con la condición de que la Significatividad sea 0.99 (99%), al obtener 2 éxitos en 10 ensayos, su Probabilidad Esperada (e) es 0.016 (1.6%) y queda muy lejos de la probabilidad de éxito de 0.20 (20%). Cuando $n = 10\ 000$ llega a $e = 0.191$ (19.1%). De $n = 10\ 000$ en adelante, el aumento de la Probabilidad Esperada (PE) es reducido. Finalmente obtenemos $e = 0.20$ (20 %) al llegar a $n = 10\ 000\ 000$. Esta característica de la Probabilidad Esperada (PE) es importante, puesto que por ella podemos saber qué probabilidad teórica hay en cada caso de 2 entre 10, 20 entre 100, 200 entre 1000, así sucesivamente. Nos llama la atención sobre todo los primeros casos donde la magnitud de la base 10, 100, 1000 es reducida, lo que causa la poca Probabilidad Esperada (0.016, 0.114, 0.171).

(3) Multiplicador

Intentamos calcular la «Frecuencia Probabilística» (FP) en forma de Probabilidad Esperada (e) * Multiplicador (m):

$$FP = e * m \quad (e: \text{Probabilidad Esperada}, m: \text{Multiplicador})$$

La Frecuencia Probabilística (FP) se obtiene por la función de la Probabilidad Esperada $E(x, n, s)$ en combinación con el Multiplicador (m).

$$FP = e * m = E(x, n, s) * m$$

$$21.5 = E(3, 3, 0.99) * 100$$

Es conveniente que la cantidad del Multiplicador (m) sea de la misma magnitud que la Máxima de Suma o Total de Palabras en forma redondeada. La precisión de la Significatividad (s) conviene ser adoptada a la magnitud del multiplicador de manera siguiente:

$$s = 1 - 1 / m$$

Por ejemplo, cuando $m = 100$, s sería $1 - 1/100 = 0.99$, y cuando $m = 100000$, s sería $1 - 1/100000 = 0.99999$.

La tabla inferior izquierda muestra la Frecuencia Absoluta (FA) y Suma vertical y la tabla derecha es de la Frecuencia Probabilística (FP) con el multiplicador (m) = 100. Ahora la Frecuencias Probabilísticas (FP) de [uoz:1200] llega a la cifra de 21.5, a diferencia de la Frecuencia Relativa (FA) de 3 entre 3: $3 / 3 * 100 = 100.0$, que es incomparable con, por ejemplo [uoz:1250], $FA = 8 / 12 * 100 = 66.7$. Resulta que la Frecuencia Relativa (FR) de 3 entre 3 (1.000) es más alta que la de 8 entre 12 (.667), mientras que la Frecuencia Probabilística (FP) de 3 entre 3 es 21.5 y la de 8 entre 12 es 30.2, lo que demuestra la importancia mayor de 8 entre 12.

FA	1200	1250	1300	1350	1400
uoz	3	8	3	11	6
boz	0	3	8	18	35
voz	0	1	1	23	53
Suma	3	12	12	52	94

FP:100	1200	1250	1300	1350	1400
uoz	21.5	30.2	3.9	9.7	1.9
boz	.0	3.9	30.2	20.0	25.9
voz	.0	.1	.1	28.2	43.9

Hemos visto que la Frecuencia Relativa (FR) no es apropiada para realizar la evaluación comparativa de las cifras. En su lugar, hemos introducido la Frecuencia Probabilística (FP) con base de la Suma de las formas en comparación. Ahora se trata de la Frecuencia Probabilística (FP) calculada con el multiplicador 100, con el que se ha utilizado la Significatividad de .99 (99%), lo suficiente para ser bastante confiable. Sin embargo, nos sorprende la magnitud de 21.5 en el caso de 3 entre 3 y la de 30.2 en el caso de 8 entre 12. Son correctos dentro de la Significatividad de .99 (99%), que es condicionada por la cantidad del Multiplicador (100). Pensamos que esto es debido a la adaptación de Suma de las frecuencias de las formas en cuestión como base de comparación. Veamos la posibilidad de la otra base, también muy usual en los estudios lingüísticos de corpus, la cantidad total de palabras o letras.

La tabla inferior izquierda es de la Frecuencia Absoluta (FA) y el número total de palabras (TP):

FA	1200	1250	1300	1350	1400
uoz	3	8	3	11	6
boz	0	3	8	18	35
voz	0	1	1	23	53
TP	7 736	36 052	40 957	64 999	96 059

FP:10^5	1200	1250	1300	1350	1400
uoz	5.6	8.1	1.1	7.3	1.9
boz	.0	1.2	7.1	14.8	23.7
voz	.0	.0	.0	20.5	39.1

La Frecuencia Normalizada (FN), que en los estudios lingüísticos de corpus suele utilizarse con el número total de palabras (TP) en forma de, por ejemplo, $3 / 7\,736 * 100\,000 = 38.8$ en [uoz: 1200] presupone que 3 entre 7 736 corresponda a 38.8 entre 100 000, por la fórmula de proporción:

$$3 / 7\,736 = 38.8 / 100\,000$$

Sin embargo, desde el punto de vista probabilístico esta presuposición no es fiable, lo mismo que la presuposición de que 3 éxitos en 10 ensayos correspondan a 30 éxitos en 100 ensayos, lo que es la base del porcentaje, como hemos visto anteriormente. En la práctica de comparación de cifras, la Frecuencia Probabilística FP es más confiable, con la que se ve un paulatino desplazamiento de <uoz> (1200-1250) por <boz> (1300) hasta <voz> (1350-1400). La misma observación es posible con la Frecuencia Probabilística (FP) de Suma con el multiplicador 100. Sin embargo, la tabla de la Frecuencia Probabilística (FP) con el número total de palabras (TP) da las cifras más realistas.

Programas

Function BinS(x, n, e)

'Significatividad (x: ocurrencia, n: ensayos, e: probabilidad expectativa)

If x = 0 Then BinS = 0: Exit Function

BinS = Application.BinomDist(x - 1, n, e, 1)

End Function

Function BinE(x, n, s) 'Probabilidad Esperada (H. Ueda 2017)

'(x: ocurrencia, n: ensayos, s: significatividad)

Dim i, k, r, mn, mx, sv: BinE = 0: k = 0: If x = 0 Then Exit Function

r = 10 ^ 6: mn = 0: mx = r

'r: precisión, mn: mínimo, mx: máximo de búsqueda binaria

Do

i = (mx + mn) / 2 'Mitad entre mx y mn

BinE = i / r 'Candidato de la Probabilidad Esperada

sc = BinS(x, n, BinE) 'sc: Significatividad del candidato

If sc < s - 1 / r Then 'Si sc no llega a s-1/r...

mx = i 'Bajar el máximo de búsqueda al punto medio (i).

ElseIf sc > s + 1 / r Then 'Si sc sobrepasa a s-1/r...

```
mn = i 'Subir el mínimo de búsqueda al punto medio (i).  
Else 'Si se encuentra en s±5/r  
Exit Do 'Salir del bucle.  
End If  
Loop  
End Function
```

En el programa (Microsoft Excel VBA) hemos utilizado la técnica de la búsqueda binaria con aplicación del encuentro en el rango. La búsqueda secuencial es impracticable por el tiempo que cuesta en llegar al encuentro.

[FIN]