LA ENTROPIA Y REDUNDANCIA DEL ESPAÑOL ESCRITO

Por: Javier Fernández G., Guillermo Gómez A., y Javier Villanueva V.*

En los trabajos a que hace referencia [1] sobre el tema de la entropía de lenguajes, notamos que ésta ha sido calculada para varios idiomas entre los cuales no se encuentra el español. El importante papel que juega este concepto en la codificación y transmisión de mensajes, análisis de textos literarios y diseño de cuestionarios nos hace pensar que ésta deficiencia debe ser corregida. El presente artículo pretende iniciar los trabajos para lograrlo. Como quedará claro en el transcurso de la exposición, los resultados aquí obtenidos no significan más que una primera aproximación cuya precisión y futuro mejoramiento sólo podrán obtenerse con el concurso de lingüistas y especialistas en muestreo.

Iniciamos el artículo con un breve resumen de los principales conceptos y resultados utilizados en él. Quien desee profundizar lo expuesto aquí puede consultar [2] y [3].

* Profesores de carrera de la Facultad de Ciencias, UNAM.

INTRODUCCION

DEFINICION 1.

Sea X una variable aleatoria (v.a.) que puede tomar los valores

$$x_1, x_2, \ldots, x_n$$

con probabilidades

$$p_1, p_2, \ldots, p_n$$

La *entropía* de X se define como

$$H(X) = -\sum_{i=1}^{n} p_i \log_2 p_i = E(-\log_2 P(X)),$$

donde

$$p_i \log_2 p_i = 0 \quad \text{si} \quad p_i = 0$$

("E" nos representa la esperanza matemática y P(X) la función de densidad de X).

Intuitivamente hablando, la entropía de un sistema (que representamos por una variable aleatoria) es una medida del grado de incertidumbre o indeterminación de éste. En caso de que interese al lector comprender cómo es que esta fórmula efectivamente refleja esa idea, le recomendamos vea la referencia [3].

La unidad de medida de la entropía es llamada *bit* y corresponde a la entropía del sistema más sencillo a plantearse, a saber, aquél que puede tomar solo dos valores con la misma probabilidad cada uno.

DEFINICION 2.

Sean X, Y variables aleatorias discretas con un número finito de valores posibles. La *entropía conjunta de* X *y* Y está dada por

$$H(X,Y) = -\sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij}\,\ell og\, p_{ij} = E(-\ell og\, P(X,Y))$$

($P(X,Y)$ función de densidad conjunta de X y Y).

La entropía de Y condicionada a que $X=x_i$ está dada por

$$H(Y|X=x_i) = -\sum_{j=1}^{m} P(Y_j|x_i)\,\ell og\, P(y_j|x_i) = E(-\ell og\, P(Y|X=x_i))$$

($P(Y|X=x_i)$ función de densidad de Y condicionada a que X toma el valor x_i).

La entropía total (completa) de Y condicionada por X se define como

$$H(Y|X) = \sum_{i=1}^{n} P_i\, H(Y|X=x_i) = E\,(-\ell og\, P(Y|X))$$

($P(Y|X)$ función de densidad de la v.a. Y condicionada por X).

PROPIEDAD 1.

Sean X,Y variables aleatorias. Entonces

$$H(Y|X) \leqslant H(Y)$$

donde la igualdad se cumple si y sólo si X y Y son independientes.

PROPIEDAD 2.

Sean X,Y variables aleatorias. Entonces

$$H(Y|X) = H(X,Y) - H(X)$$

Sean X y Y variables aleatorias. La *cantidad de información promedio* sobre X contenida en Y(Notación $I_{Y \to X}$) es la entropía perdida por X una vez que Y queda determinada, i.e.

$$I_{Y \to X} = H(X) - H(X \mid Y).$$

En general, la cantidad de información I_X contenida en una variable aleatoria X, es decir, la cantidad de información obtenida al conocer el valor que toma la v.a. X está determinada por su entropía, esto es $I = H(X)$

1. ENTROPIA DEL ESPAÑOL ESCRITO

Si todas las letras en nuestro lenguaje se presentaran con la misma frecuencia, tendríamos que en base a la definición 1, la entropía de una letra del mismo estaría dada por

$$H_0 = \log 28 = 4.80735 \text{ bits.} \quad *$$

Sin embargo, ésta es una condición que evidentemente no satisface ninguna lengua. Para calcular más exactamente la cantidad de información contenida en una letra del alfabeto castellano necesitamos conocer las frecuencias relativas de las diferentes letras. -

* Estamos considerando como una letra más el espacio entre palabras (letra neutral).

Ellas pueden ser calculadas en forma aproximada tomando textos representativos sobre diferentes temas del español escrito. Nosotros utilizamos una muestra de cerca de 80000 letras, que comprende textos científicos, de humanidades y de publicaciones periódicas en diferentes proporciones. En el suplemento aparecen los resultados obtenidos, así como la naturaleza de la muestra en detalle.

Utilizando estas frecuencias tendríamos que la entropía H_1 de de una letra del alfabeto castellano, estaría dada por (de acuerdo con la definición 1)

$$H_1 = H(\alpha_1) = -\sum_{i=1}^{28} fr(i) \ \ell og \ fr(i) = 3.99309 \text{ bits},$$

donde α_1 = "experimento que permite determinar una letra con la -- frecuencia que aparece en el castellano escrito", y $fr(i)$ = frecuencia relativa de la letra i-ésima.

Esto nos lleva a una disminución igual a $H_0 - H_1 = .81426$ bits en la cantidad de información contenida en una letra.

Ahora bien; el calcular $H_1 = H(\alpha_1)$, se considera que las letras aparecen en forma independiente, es decir, que la aparición de una letra no depende de cuál haya sido la anterior. Sin embargo, esta restricción es evidentemente muy fuerte, pues basta un análisis superficial para observar que por ejemplo después de una consonante aumenta grandemente la probabilidad de que la siguiente letra sea vocal, o que después de la **q** en textos literarios sólo puede escribirse **u**, etc., es decir, que las letras no aparecen en forma independiente. Se ocurre entonces calcular la entropía contenida en el experimento que consiste en determinar una letra del alfabeto castellano bajo la condición de que conocemos la letra que le antecede, es decir, calcular

$$H_2 = H(\alpha_2|\alpha_1) = H(\alpha_1 \cap \alpha_2) - H(\alpha_1) \; *$$

(donde $\alpha_1 \cap \alpha_2$ = experimento consistente en determinar dos letras con la frecuencia que aparecen simultáneamente en el español escrito), para lo cual es necesario conocer las frecuencias de todas las posibles parejas de letras (bigramas) que pueden aparecer en un texto escrito.

Esto fue hecho utilizando la misma muestra mencionada anteriormente, y se obtuvo que

$$H(\alpha_1 \cap \alpha_2) = 7.16038 \text{ bits}$$

de donde

$$H_2 = 3.16729 \text{ bits}$$

reduciéndose así la entropía de una letra en

$$H_1 - H_2 = 0.82580 \text{ bits}$$

Sin embargo, la dependencia de ocurrencia de una letra no es sólo respecto a la inmediata anterior, sino a varias de las anteriores; por ejemplo, después de la "qu" es muy frecuente que ocurra "e", pero poco frecuente que aparezca una consonante; antes de la "a" es más frecuente la combinación "br" que la "aa", etc. Así pues, debemos encontrar

$$H_3 = H(\alpha_3|\alpha_1 \cap \alpha_2) = H(\alpha_1 \cap \alpha_2 \cap \alpha_3) - H(\alpha_1 \cap \alpha_2)$$

* Proposición 2.

para lo cual es necesario conocer las frecuencias de todas las posibles combinaciones de tres letras del español escrito (trigramas); en base a los resultados obtenidos, llegamos a que

$$H_3 = 2.66265 \text{ bits},$$

reduciéndose la entropía de una letra en $H_2 - H_3 = 0.50464$ bits.

En castellano como en otras lenguas, es claro que, la aparición de una letra no depende sólo de cuales fueron las 1 ó 2 anteriores, sino de muchas más.

De esta forma, podemos decir que las entropías

$$\begin{aligned} H_N &= H(\alpha_N | \alpha_1 \cap \alpha_2 \cap \ldots \cap \alpha_{N-1}) \\ &= H(\alpha_1 \cap \alpha_2 \cap \ldots \cap \alpha_N) - H(\alpha_1 \cap \alpha_2 \cap \ldots \cap \alpha_{N-1}), \end{aligned}$$

$$N = 0,1,2,3,\ldots$$

no son sino aproximaciones a la entropía de este sistema, es decir, a la entropía del lenguaje, y es evidente que conforme mayor es la N, mejor es la aproximación a la entropia buscada. Esto quedaría formalmente justificado haciendo notar que al crecer la N la H_N correspondiente sólo puede decrecer*, y como $H_N > 0$ entonces existe

$$H = \lim_{N \to \infty} H_N$$

Es este valor de H el que nos determina la *entropía del español* (y en general de cualquier lenguaje).

* En efecto, puede demostrarse rigurosamente que $H(\beta | \alpha \cap \gamma) \leq H(\beta | \gamma)$ utilizando la propiedad 1.

Desde 1951 Shannon calculó para el idioma inglés H_2 y H_3, e hizo aproximaciones para H_5 y H_8, utilizando las frecuencias de -ocurrencia de palabras diferentes. La idea es la siguiente:

Conocidas las frecuencias de ocurrencia $p_1,\ldots,p_K$ de las K palabras diferentes del español escrito, definiremos la "*entropía de palabra de 1er. orden*" como

$$H_1^{(P)} = - \sum_{i=1}^{K} p_i \,\ell og\, p_i$$

Ahora bien; si w es el promedio de letras por palabra, entonces el cociente

$$\frac{H_1^{(P)}}{w}$$

nos dá una estimación de la entropía condicional de orden w, es decir, de H_w. En efecto:

$$\frac{H_1^{(P)}}{w} < H_w$$

ya que la relación de dependencia entre w letras de una misma palabra es notablemente mayor que la relación de dependencia de sucesiones de w letras arbitrariamente tomadas de un texto entendible. Y, por otro lado,

$$H < \frac{H_1^{(P)}}{w}$$

ya que la magnitud $H_1^{(P)}$ no contempla en absoluto la dependencia existente entre las palabras de un texto dado.

En base a nuestra muestra, nosotros obtuvimos que, en castellano:

$$H_1^{(P)} = 9.32665 \text{ bits},$$

y habiendo también obtenido que el número de letras que en promedio tiene una palabra es 5.566, llegamos a que

$$H_6 \approx \frac{H_1^{(P)}}{6} = 1.55442 \text{ bits}.$$

Sin embargo, éste valor no es aún una buena aproximación al valor real de H, pues presupone que las palabras en castellano aparecen en forma independiente; y ésto es claro que no es real: por ej., después de un pronombre es muy frecuente la aparición de un verbo, no así la de otro pronombre, etc. Para obtener una mejor aproximación resulta necesario calcular la *entropía de palabra de 2º orden*, que estará dada por

$$H_2^{(P)} = - \sum_{i=1}^{s} r_i \log r_i$$

donde $r_i (i=1,2,\dots,s)$ nos representan las frecuencias de todas las s posibles parejas de palabras (bipalabras). De manera que si b nos denota la longitud promedio de una bipalabra, a través de razonamientos equivalentes a los utilizados anteriormente se prueba que

$$H < \frac{H_2^{(P)}}{b} < H_b$$

donde H_b es la *entropía condicional de orden* b.

Habiendo calculado las frecuencias de ocurrencia de todas las posibles bipalabras, nosotros obtuvimos que, para el castellano escrito

$$H_2^{(P)} = 13.00506$$

y como la longitud promedio de bipalabra obtenida fué de **10,9222**, resulta que

$$H_{11} \approx \frac{H_2^{(P)}}{11} = 1.182278 \text{ bits.}$$

Sin embargo, como ya vimos anteriormente, éstas no son sino aproximaciones a la entropía del lenguaje. Lo que no sabemos, es qué tan buenas son estas aproximaciones. Es más seguirnos aproximando a **H** en forma directa resulta prácticamente imposible (p.ej., para calcular H_{12} necesitaríamos saber las frecuencias de ocurrencia de 12 letras, que son del orden de trillones (28^{12})).

A Shannon se debe la idea de utilizar métodos indirectos para calcular H_N para **N** grande. Considerando a H_N como la entropía condicional de orden **N**, es decir, el grado de incertidumbre (o indeterminación) del experimento α_N consistente en determinar la **N**-ésima letra de un texto bajo la hipótesis de que conocemos las **N-1** anteriores, la idea de Shannon es la siguiente:

Escogemos aleatoriamente un segmento de **(N-1)** letras de un texto, y un interlocutor debe predecir cual es la **N**-ésima; en particular, se propone preguntar a diferentes personas y tomar aquella - respuesta que sea la más acertada, ya que la predicción se supone que se hace de la manera más racional posible, es decir, con conocimiento de las regularidades estadísticas propias del idioma. Este experimento puede repetirse un número adecuado de veces y la dificultad de predecir la N-ésima letra puede estimarse mediante el promedio Q_N del número de pruebas necesarias para una respuesta correcta. Es claro que este promedio de pruebas sólo puede decrecer al crecer N. Si a partir de una cierta N ya no decrece Q_N, esto va a significar que los experimentos correspondientes tienen el mismo grado de incertidumbre (indeterminación), es decir, que la - entropía correspondiente H_N prácticamente no difiere de su valor - límite H.

Siguiendo estos razonamientos Shannon efectuó una serie de experimentos para el idioma inglés con $N = 1,2,3,4,\ldots,13,15$ y 100, y observó que la predicción de la centésima letra a partir de las 99 anteriores resultaba mucho más sencilla que predecir la quinceava conociendo las 14 anteriores, lo cual indica que H_{15} es sensiblemente mayor que H, ésto es, H_{15} es mala aproximación a H. Experimentos más refinados fueron realizados por Burton, Licklider para $N = 1,2,4,8,16,32,64,128$ y $N \simeq 10000$, llegando al siguiente resultado: H_{32}, H_{64} y H_{128} "prácticamente" no difieren de H_{10000}. - Sin embargo, H_{16} es significativamente mayor que esta magnitud; - por esto se toma H_{30} ó H_{40} como H.

Lo anterior indica que nuestra mejor aproximación a la entropía del español escrito (H_{11}) es aún una mala aproximación. Sin em bargo, hay una pequeña diferencia con respecto al método utilizado por Shannon que posiblemente se refleja en los valores obtenidos; y ella es que, por un lado, la muestra por nosotros analizada es considerablemente mayor, y por otro, que los cálculos de las frecuencias correspondientes fueron hechos con un método más preciso [3]. Ello nos hace pensar que nuestras aproximaciones (en particular H_{11}) deberán ser mucho más cercanas al valor real de H que las correspondientes de Shannon para el inglés.

Esto se ve corroborado al comparar los resultados de Shannon para el idioma inglés con los obtenidos por nosotros para el español (y es más claro) si comparamos la rapidez con que H_N decrece en ambos casos, lo cual se ve reflejado en las diferencias):

VALORES DE H_N OBTENIDOS POR SHANNON PARA EL INGLES	VALORES DE H_N OBTENIDOS POR NOSOTROS PARA EL ESPAÑOL
$H_0 = 4.76$	$H_0 = 4.81$
$H_1 = 4.03$	$H_1 = 3.99$
$H_2 = 3.32$	$H_2 = 3.17$
$H_3 = 3.10$	$H_3 = 2.66$
$H_5 = 2.10$	$H_6 = 1.55$
$H_8 = 1.90$	$H_{11} = 1.18$
$0.6 < H_{100} < 1.3$	

DIFERENCIAS	DIFERENCIAS
$H_0 - H_1 = 0.73$	$H_0 - H_1 = 0.82$
$H_1 - H_2 = 0.71$	$H_1 - H_2 = 0.82$
$H_2 - H_3 = 0.22$	$H_2 - H_3 = 0.51$

Observemos que la sucesión $\{H_N\}$ de la derecha decrece más rápidamente que la de la izquierda; de ahí que sea de esperarse que la **N** a partir de la cual los valores de H_N se consideran una buena aproximación a su valor límite **H**, sea menor para la sucesión de la derecha que para la de la izquierda.

Sin embargo, como decíamos al principio de este artículo, nuestro trabajo no pretende sino ser un primer intento en la investigación de este tipo de características del lenguaje, que deberá ser mejorado con la participación de gente especializada en el área.

2. REDUNDANCIA DEL ESPAÑOL ESCRITO

La información promedio **H** contenida en un símbolo de una lengua es considerablemente menor que su valor máximo posible H_0 (ello es evidente de la primera sección. Esto significa que con frecuencia partes de un mensaje dicen cosas que parcialmente ya se sabían. Una medida de esta parcial o completa repetición del contenido de un mensaje que ocurre en una lengua es la *redundancia*. La redundancia **R** de un lenguaje se define por

$$R = 1 - \frac{H}{H_0}$$

y nos representa (burdamente) el porcentaje con que queda definida la siguiente letra de un texto, por la estructura misma de la lengua en cuestión (a diferencia de quedar definida aleatoriamente*). Mediante la codificación de mensajes, de manera que sea eliminada la redundancia, es posible incrementar la información por símbolo. Sin embargo, a pesar de la pérdida que ocasiona en la transmisión de información, la redundancia es una propiedad muy útil de los lenguajes que permite reconocer errores individuales en la transmisión de mensajes de manera que puedan ser fácilmente corregidos. Digamos que la redundancia es el mecanismo "corrector de errores" de los lenguajes. Así pues, si redujésemos los textos telegráficos mediante la cancelación de ciertas palabras, la redundancia del idioma nos permitiría reconstruirlos; es la redundancia la que nos permite entender el texto de un libro no obstante los errores del mismo.

Supongamos, para aclarar un poco más el sentido de **R**, que el idioma castellano es codificado con un alfabeto código de 28 caracteres (absurdo y artificial pero es útil suponerlo para entender mejor el significado de **R**). En caso de codificación óptima, para escribir (transmitir) un mensaje de **M** letras (con **M** suficientemente grande) necesitamos en promedio

$$\frac{H}{H_0} M = (1-R)M = M-RM \text{ caracteres código; } **$$

* Obsérvese que si quedase definida aleatoriamente por la estructura de la lengua la siguiente letra de un texto, **R** en base a la anterior interpretación, debiera ser cero. Y en efecto, porque si tal fuese el caso tendríamos que $H=H_0$ y por tanto $R = 0$.

** Un mensaje de M símbolos (M suficientemente grande) tomados de un sistema X y codificado a través de un alfabeto de m caracteres tiene una longitud no menor de $M \frac{H(X)}{\log} \log m$ caracteres código. (Ver [3]) es decir, respecto a la escritura usual podemos simplificar el texto en RM letras. Esto no significa que podemos arbitrariamente eliminar RM letras y del resto restablecer (decodificar) sin errores - el texto original; para poder hacer esta simplificación es necesario utilizar el mejor método de codificación, tal que luego de su aplicación obtengamos independencia y equiprobabilidad de todas las letras del mensaje. A priori podemos afirmar que si despreciamos más de - RM letras con toda seguridad no podremos decodificar sin errores el texto original.

La redundancia resulta ser una característica muy importante - de un idioma; sin embargo, su valor numérico no ha sido establecido satisfactoriamente para ninguna lengua. En castellano a la fecha no se tenían ni siquiera estudios que estableciesen las magnitudes de H_2 y H_3. En base a los valores que nosotros hemos encontrado podemos al menos afirmar que para nuestro idioma

$$R \geqslant 1 - \frac{H_{11}}{H_0} \simeq 1 - .2459414 = 0.7540687$$

Es decir, que el castellano escrito es en más de un 75% redundante.

SUPLEMENTO

Para calcular la entropía y la redundancia del castellano, así como para codificarlo, fué necesario estudiar datos referentes al - lenguaje, tales como con qué frecuencia aparecen cada una de las letras, bigramas y trigramas, palabras y parejas de palabras (bipalabras) en el español escrito.

Para ello extrajimos una muestra que comprende diferentes áreas de textos científicos, de humanidades y publicaciones periódicas. - Las áreas consideradas y las proporciones correspondientes están dadas por el siguiente esquema:

			(Porcentaje relativo a la rama)
CIENCIAS 25% (del total)	Física	25%	"
	Química	25%	"
	Matemáticas	25%	"
	Biología	25%	"
HUMANIDADES 25% (del total)	Letras	20%	"
	Filosofía	15%	"
	Historia	15%	"
	Psicología	15%	"
	Economía	20%	"
	Derecho	15%	"
PUBLICACIONES PERIODICAS 50% (del total)	Periódicos	50%	"
	Revistas	25%	"
	Cuentos	25%	"

El tamaño de la muestra utilizada es de 80,490 letras.

A continuación presentamos algunos de los resultados más relevantes obtenidos en los diferentes programas (cuyos listados y resultados íntegros aparecen en [3]).

FRECUENCIA DE LAS LETRAS DEL ESPAÑOL ESCRITO*

—	.168604**	R	.052888	P	.021145	F	.005988
E	.114225	L	.046055	B	.009641	Z	.003143
A	.098112	D	.041384	G	.009069	J	.002845
O	.077513	C	.038725	V	.008473	X	.001876
S	.065238	T	.038166	Q	-007914	Ñ	.001006
N	.058566	U	.034911	Y	.006436	K	.000112
I	.056541	M	.025171	H	.006137	W	.000062

BIGRAMAS CON UNA FRECUENCIA MAYOR QUE 10^{-2}

A—	3.183×10^{-2}	ES	1.730×10^{-2}	ER	1.165×10^{-2}
E—	3.179×10^{-2}	—L	1.559×10^{-2}	AS	1.116×10^{-2}
S—	3.112×10^{-2}	OS	1.459×10^{-2}	NT	1.085×10^{-2}
D—	2.555×10^{-2}	—C	1.366×10^{-2}	RA	1.080×10^{-2}
—E	2.061×10^{-2}	LA	1.340×10^{-2}	L—	1.067×10^{-2}
EN	1.954×10^{-2}	—A	1.293×10^{-2}	AR	1.033×10^{-2}
DE	1.925×10^{-2}	—P	1.280×10^{-2}	RE	1.015×10^{-2}
—J	1.822×10^{-2}	—S	1.227×10^{-2}	EL	1.005×10^{-2}
N—	1.782×10^{-2}	ON	1.165×10^{-2}		

* Todas las frecuencias aparecen en orden decreciente

** El símbolo "—" significa espacio entre palabras.

TRIGRAMAS CON UNA FRECUENCIA MAYOR QUE 5×10^{-3}

-DE	1.577×10^{-2}	ENT	6.522×10^{-3}
OS-	1.290×10^{-2}	QUE	6.348×10^{-3}
DE-	1.172×10^{-2}	-QU	6.323×10^{-3}
-LA	9.293×10^{-3}	-EL	6.125×10^{-3}
AS-	8.771×10^{-3}	UE-	6.038×10^{-3}
ES-	8.212×10^{-3}	-EN	5.926×10^{-3}
LA-	7.404×10^{-3}	ON-	5.876×10^{-3}
EL-	7.206×10^{-3}	-ES	5.479×10^{-3}
EN-	6.659×10^{-3}	-SE	5.180×10^{-3}

PALABRAS CON UNA FRECUENCIA MAYOR QUE 5×10^{-3}

DE	6.465×10^{-2}	LAS	1.298×10^{-2}
LA	3.941×10^{-2}	SUS	1.151×10^{-2}
QUE	3.225×10^{-2}	NO	1.077×10^{-2}
EL	3.107×10^{-2}	POR	1.055×10^{-2}
EN	2.693×10^{-2}	DEL	9.594×10^{-3}
YA	2.442×10^{-2}	ES	9.594×10^{-3}
LOS	2.295×10^{-2}	CON	8.192×10^{-3}
UNA	2.088×10^{-2}	PARA	6.568×10^{-3}
ABSOLUTO	2.073×10^{-2}	ALIMENTA	6.494×10^{-3}
SESENTA	1.668×10^{-2}		

BIPALABRAS CON UNA FRECUENCIA MAYOR QUE 10^{-3}

DE LA	9.732×10^{-3}	DE QUE	1.426×10^{-3}
A LA	4.363×10^{-3}	QUE NO	1.258×10^{-3}
DE LOS	4.363×10^{-3}	EN LAS	1.174×10^{-3}
EN EL	3.775×10^{-3}	Y EL	1.174×10^{-3}
EN LA	3.440×10^{-3}	CON EL	1.090×10^{-3}
DE LAS	2.685×10^{-3}	EN UN	1.090×10^{-3}
DE UNA	2.517×10^{-3}	EN LOS	1.090×10^{-3}
QUE SE	2.349×10^{-3}	A UN	1.006×10^{-3}
QUE EL	1.762×10^{-3}	LO QUE	1.006×10^{-3}
A LOS	1.594×10^{-3}	Y ES	1.006×10^{-3}
DE SU	1.510×10^{-3}	EN SUS	1.006×10^{-3}

BIBLIOGRAFIA

[1] Yaglom, A.M., Yaglom I. M., *"Veroyatnost y Informatzii"* Nauka, Moscú, 1973.

[2] ASH R.; *Information theory*, Interscience Pub., John Wiley, New York, 1965.

[3] Fernández G. J.; *Acerca de la teoría de información y algunas de sus aplicaciones*, Tesis Profesional, Facultad de Ciencias, UNAM, 1975.