Decryption Through the Likelihood of Frequency of Letters

Barbara Sánchez Rinza, Fernando Zacarias Flores, Luna Pérez Mauricio, and
Martínez Cortés Marco Antonio

Benemérita Universidad Autónoma de Puebla,
Computer Science
14 Sur y Av. San Claudio, Puebla, Pue.
72000 México
brinza@cs.buap.mx, fzflores@yahoo.com.mx

Abstract. The method to decrypt the information using probability
leads to a more thorough job, because you have to know the percent-
age of each of the letters of the language that is being analyzed here is
Spanish. You can consider not only the probabilities of the letters also
syllables, set of three, four letters and even words. Then you have this
thing to do is make comparisons of the frequencies of cipher text and
the frequencies of the language to begin to replace by a correspondence.
And finally passing a scanner and find the decrypted text.

Keywords Probability, Decrypt.

1 Introduction

Cryptography is the science that alters the linguistic representations of a message
[1]. For this there are different methods, where the most common is encryption.
This science masking the original references of the information by a conversion
method governed by an algorithm that allows the reverse or decryption of in-
formation. Use of this or other techniques, allowing for an exchange of messages
that can only be read by the intended beneficiaries as 'consistent'. A consistent
recipient is the person to whom the message is directed with the intention of
the sender. Thus, the recipient knows the discrete coherent used for masking the
message. So either have the means to bring the message to the reverse process
cryptographic, or can infer the process that becomes a message to the public. The
original information to be protected is called plaintext or cleartext. Encryption
is the process of converting plain text into unreadable gibberish called cipher-
text or cryptogram. In general, the concrete implementation of the encryption
algorithm (also called figure) is based on the existence of key secret information
that fits the encryption algorithm for each different use [2].

Decryption is the reverse process to recover the plaintext from the ciphertext
and key. Cryptographic protocol specifies the details of how to use algorithms
and keys (and other primitive operations) to achieve the desired effect. The set

of protocols, encryption algorithms, key management processes and actions of the users, which together constitute a cryptosystem, which is what the end user works and interacts. In this work, we must first have a ciphertext which must meet certain requirements, such a text should be bijective so that each element of the domain carries a single element of the condominium. In addition we must also take account of the rules of Kerckhoff [3].

2 Development work

2.1 Frequencies in Spanish

Is required to decrypt text using the odds as to how often they used certain letters in the alphabet, for this work only considered the Spanish language [5].

The frequencies of Spanish, which were used for this study were:

1. *Frequency triglyphs*
2. *Frequency of digraphs*
3. *Most common words*
4. *Frequency of letters at the beginning of words*
5. *Frequency of letters in Spanish*
6. *Frequency Words*

2.2 Triglyphs Frequencies

The letter frequency statistics may vary from one to another depending on the corpus author has chosen to develop them. Usually differences when the corpus is literary or consists of texts of different origins. Table 1 shows the frequency of each of the Spanish alphabet with their respective percentage.

High frequency letters		Medium frequency letters		Low frequency letters		Frequencies 0.5%
letter	freq.%	letter	freq.%	letter	freq.%	G, F, V, W
E	16,78	R	4,94	Y	1,54	
A	11,96	U	4,80	Q	1,53	
O	8,69	I	4,15	B	0,92	
L	8,37	T	3,31	H	0,89	
S	7,88	C	2,92			J, Z, X, K, N
N	7,01	P	2,76			
D	6,87	M	2,12			

Table 1. Frequency triglyphs

2.3 Most Frequent words

The vowels make up about 46.38% of the text. The high frequency letters account for 67.56% of the text. Mid-frequency points accounting for 25% of the text [4]. In the dictionary the most common vowel is A, but in written texts is the E because of prepositions, conjunctions, verbs, etc. The most common consonants are L, S, N, D, with about 30%. The less frequent six letters: V, N, J, Z, X and K (just over 1%). The average frequency of a Spanish word is 5.9 letters. The coincidence index for Spanish is 0.0775. In addition to solving the encryption table 2 we mentioned that we most frequently used words in a text of 10 000 words.

Most common words		Two-letter words	Three-letter words	
Word	Frequency	Frequency	Word	Frequency
DE	778	778	QUE	289
LA	460	460	LOS	196
El	339	339	DEL	156
EN	302	302	LAS	114
QUE	289	119	POR	110
Y	226	98	CON	82
A	213	74	UNA	78
LOS	196	64	MAS	36
DEL	156	63	SUS	27
SE	119	47	HAN	19
LAS	114			

Table 2. Most frequent words of one, two and three letter

Next, table 3 shows the frequencies of the 4-letter words.

2.4 Frequency digraphs

The size of the corpus is 60,115 letters. The frequencies are absolute. The digraphs are read by row and column in that order. Below in table 4 shows the union digraphs are letters from letters.

2.5 Most common initial letter

The most frequent letters in Spanish that start a word are listed in Table 5

3 Results

The ciphertext is used as said it had to be bijective and have Kerckhoff rules and the decrypted text shown in Figure 1.

Four-letter words		Distribution of letters in literary texts			
Word	Frequency	E - 16,78%	R - 4,94%	Y - 1,54%	J - 0,30%
PARA	67	A - 11,96%	U - 4,80%	Q - 1,53%	
COMO	36	O - 8,69%	I - 4,15%	B - 0,92%	
AYER	25	L - 8,37%	T - 3,31%	H - 0,89%	
ESTE	23	S - 7,88%	C - 2,92%	G - 0,73%	
PERO	18	N - 7,01%	P - 2,77%	F - 0,52%	
ESTA	17	D - 6,87%	M - 2,12%	V - 0,39%	
AOS	14				
TODO	11				
SIDO	11				
SOLO	10				

Table 3. Frequency with four letters

4 Conclusions

We conclude that this method of decryption is good however would have to tweak a little more due to it depends on the text we have and how much text to decrypt was also observed that only decrypts an encrypted bijective. In this work, as seen in the results of Figure 1, which apply various processes, first see the probability of the lyrics in Spanish that are more frequent, then seen with the syllables that are more frequent in Spanish, and then with the last word and you miss the information, text analyzer, as shown in Figure 1 a large percentage of the information is decoded, but as mentioned in the top, this will depend have that much information to process it.

References

1. Liddell and Scott's Greek-English Lexicon. Oxford University Press. (1984)
2. Anaya Multimedia, Codigos Y Claves Secretas: Programas En Basic, Basado A Su Vez En Un Estudio Lexicogrfico Del Diario "El Pas", Mexico 1986.
3. Friedman, William F. And Callimahos, Lambros D., Military Cryptanalytics, Cryptographic Series, 1962
4. Part I - Volume 2, Aegean Park Press, Laguna Hills, Ca, 1985
5. Barker, Wayne G., Cryptograms In Spanish, Aegean Park Press, Laguna Hills, Ca.,

	A	B	C	D	E	F	G	H	I	J	K	L	M
A	12	14	54	64	15	5	8	4	10	8		41	30
B	11				5				14	1		12	
C	39		5		17			8	80			3	
D	32		1	2	84			1	30				
E	20	5	47	26	17	8	21	6	9	3		44	26
F	2				9				12			1	
G	12				12				5			1	
H	15				3				5				
I	43	8	42	29	40	5	8			1		14	16
J	4				5								
K					1								
L	44		5	5	35	1	3		28			9	5
M	32	10			42				30				
N	41	2	33	37	41	10	6	2	28	1		5	4
O	19	17	28	26	16	6	5	5	4	1		22	33
P	30		1		16				5			8	
Q													
R	74	1	12	10	94	1	12		45	1	1	6	15
S	32	2	18	15	57	3	2	4	41	1		5	7
T	60		1		67				35				
U	13	6	11	5	52	1	3		9			9	6
V	12			1	15				15				
W	1				1								
X			1		4								
Y	5	1	3	2	5	1	1					1	1

Table 4. Frequency of digraphs

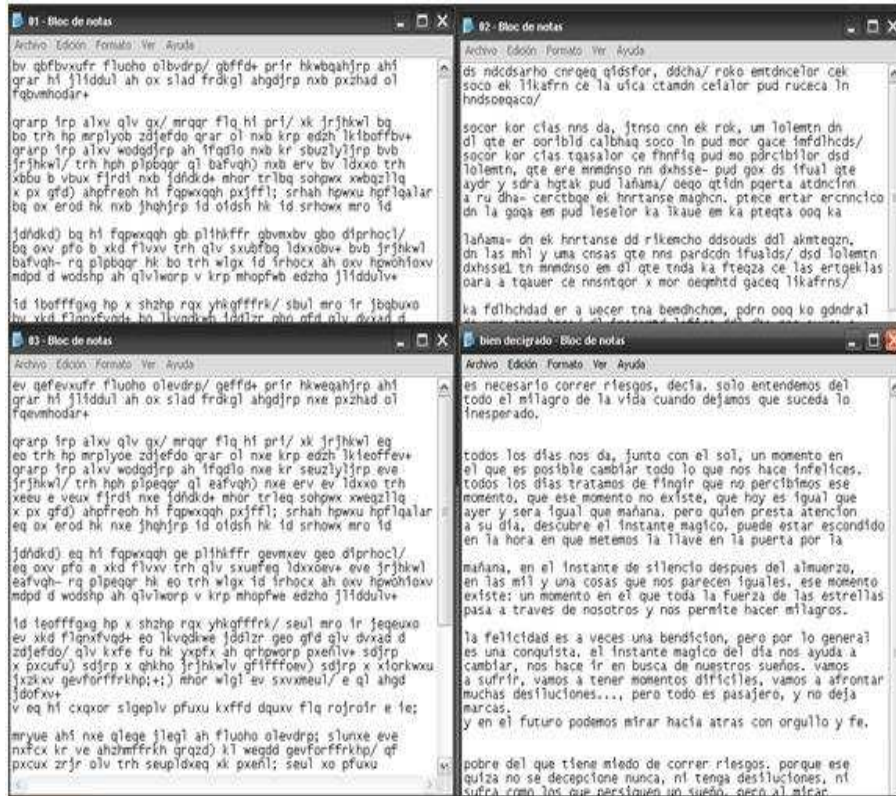| letter	P	C	D	E	S	A	L	R	M	N	T			
frequency	1.1128	1.081	1.012	989	789	761	435	425	403	346	298			
letter	Q	I	H	U	G	V	F	O	B	J	Y	W	Z	K
frequency	286	281	230	219	206	183	177	169	124	47	27	19	2	1

Table 5. Frequency of initial letters

Fig. 1. with each of the texts worked, 01 encrypted text, 02 text one pass, 03 second pass the text, either original text decrypted