ATTACKING TURKISH TEXTS ENCRYPTED BY HOMOPHONIC CIPHER

(HOMOFONİK ŞİFRELENMİŞ TÜRKÇE METİNLER ÜZERİNE ATAKLAR)

by

Şefik İlkin SERENGİL, B.S.

Thesis

Submitted in Partial Fulfillment

of the Requirements

for the Degree of

MASTER OF SCIENCE

in

COMPUTER ENGINEERING

in the

INSTITUTE OF SCIENCE AND ENGINEERING

of

GALATASARAY UNIVERSITY

June 2011

ATTACKING TURKISH TEXTS ENCRYPTED BY HOMOPHONIC CIPHER

(HOMOFONİK ŞİFRELENMİŞ TÜRKÇE METİNLER ÜZERİNE ATAKLAR)

by

Şefik İlkin SERENGİL, B.S.

Thesis

Submitted in Partial Fulfillment

of the Requirements

for the Degree of

MASTER OF SCIENCE

Date of Submission : May 18, 2011

Date of Defense Examination: June 17, 2011

Supervisor : Asst. Prof. Dr. Murat AKIN

Committee Members : Assoc. Prof. Dr. Tankut ACARMAN

Assoc. Prof. Dr. Y. Esra ALBAYRAK

Acknowledgements

I would like to thank with my whole heart to my supervisor, Asst. Prof. Dr. Murat Akın, who taught me a lot of things in my academic life.

I am grateful to my beloved mother and father who do not avoid sacrificing anything from the day I was born to ensure me to have the best conditions.

May, 2011
Şefik İlkin Serengil

Table of Contents

List of Figures

List of Tables

Abstract

Emerging technologies make the vital operations performing through insecure channels. At this point, transferring private information between parties and authentication of transferred data can only be possible by the use of cryptography.

The security of the cryptographic methods should be examined to test for perfection. Therefore, it is always supposed that the cryptographic method is known on cryptographic attacks.

In this work, firstly classical encryption methods and vulnerabilities for Turkish language are reviewed. Herein, homophonic cipher comes one step forward in the alternative classical encryption methods because it generates ciphertexts consisting of variable block sizes. This makes well known attacking models invalid. Therefore, a novel attacking model is aimed to develop for Homophonic cipher in Turkish. This model is investigated on a large data source aiming to detect the characteristic features of Turkish language.

Key words: Turkish n-gram Frequencies, Homophonic Substitution, Cryptoanalysis of Encrypted Turkish Texts.

Résumé

Le développement de la technologie offre la possibilité d'effectuer des opérations vitales sur des canaux dont la sécurité n'est pas sûre. A ce point-là, le transfert de l'information privée entre les parties et l'authentification de ces données transférées sont possibles seulement grâce à l'utilisation de la cryptographie.

Pour pouvoir examiner la perfection d'une méthode cryptographique, il faut d'abord questionner la sécurité qu'il offre à l'utilisateur. Par conséquence, on suppose toujours que la méthode du chiffrement est connue quand on parle d'une attaque cryptographique.

Dans ce travail, premièrement on révise les méthodes de chiffrement classique et on parle des côtés vulnérables de la langue turque. Le chiffrement homophonique se diffère des autres méthodes de chiffrement grâce à sa particularité qui rend possible de produire des chiffres de blocs de taille variable. Cette caractéristique rend inefficaces les attaques effectuées en utilisant des méthodes bien connues. Pour cette raison, le but de ce travail est de proposer un nouveau modèle d'attaque propre aux textes turcs cryptés par le chiffrement homophonique. Le modèle est examiné sur une grande base de données afin d'évaluer les caractéristiques de la langue turque.

Mots clés: Les fréquences des n-gram turcs, la substitution homophonique, cryptanalyse des textes turcs.

Özet

Gelişen teknoloji ile güvenli olmayan kanallardan hayati işlemlerin gerçekleştirilmesi günümüz dünyasında oldukça yaygın bir şekilde yer almaktadır. Bu noktada, gizli bir bilginin partiler arasında aktarılması ve aktarılan bilgilerin doğruluğunun taahhüt edilmesi ancak ancak kriptografi ile mümkün olmaktadır.

Bir kriptografik yöntemin mükemmelliğinin sınanabilmesi için güvenliği sorgulanabilmelidir. Bu nedenle, kriptografik ataklarda hep kriptografik yöntemlerin bilindikleri varsayılır.

Bu çalışmada öncelikle kriptografinin temelini oluşturan klasik yöntemler ve Türkçe'ye özgü zafiyetlere değinilmiştir. Bu noktada homofonik şifreleme, değişken blok boyutunda şifreler üretmesi sebebiyle alternatif şifreleme yöntemleri arasında kendini belli etmektedir. Bu da bilinen yöntemler ile saldırıları etkisiz kılmaktadır. Bu nedenle, homofonik şifrelenmiş Türkçe metinler için özgün bir saldırı modeli geliştirilmesi amaçlanmıştır. Bunun için oldukça geniş bir veri kaynağı üzerinde inceleme yapılmış ve Türkçe'nin karakteristik özellikleri kestirilmeye çalışılmıştır.

Anahtar Sözcükler: Türkçe n-gram Frekansları, Homofonik Yer Değiştirme, Türkçe Şifrelenmiş Metinlerin Kriptoanalizi.

1. Introduction

Cryptography could be implemented to provide data security, integrity and authentication. Moreover, the cryptosystems have to be resistant against to attacks. Therefore, detection of the vulnerebalities of a cryptographic method is a valueable processus. Hence, it is aimed to mention the main concepts of secret key cryptography and it is planned to focus on the attacking approaches of the secret key systems in this work.

Statistical features of the source language are used to solve classical methods because classical methods depend on the source language. Related work presented by Dalkılıç [1] investigates Turkish language patterns and frequencies of Turkish. This contribution solves some of classical methods but this study is limited by the analysis about extraction of most frequent trigrams.

Classical cryptography is presented in the first section. The main concepts of well known classical encryption methods are considered and attacking models are illustrated.

Most particularly, there seems to have been no previous work on the subject that analyses homophonic cipher for Turkish. Therefore, developing an attacking model for Turkish is notably aimed in the second section.

In this work, the corpus of size 13.4 MB is used to attain the statistical features of Turkish to solve some of classical cryptographic methods. Also, the corpus consists of 120 articles of a columnist, *Çetin Altan,* from the Turkish daily newspaper Milliyet and 37 novels of 9 different authors, which are *Orhan Kemal, Orhan Pamuk, Çetin Altan, Aziz Nesin, Rıfat Ilgaz, Gülse Birsel, Ahmet Altan, Yılmaz Erdoğan* and *Soner Yalçın*.

2. Classical Cryptography

Classical cipher is an encryption method that performs on letters of alphabet. Classical ciphers depend on the source language. Therefore, source language specifies the key space of the method. More generally, they are implemented by hand or they are implemented with basic machines. Substitution or transposition techniques are often included in classical ciphers. In this section, most popular methods of classical cryptography and main concepts are presented.

2.1. Monoalphabetic Substitution

Monoalphabetic substitution is a historical encryption method that depends on replacing each plaintext letter with another letter. The cipher alphabet remains unchanged throughout the encryption process. Therefore, it is named as *Monoalphabetic Substitution Cipher*.

2.1.1. Shift Cipher

Shift cipher depends on the principle that each character of the message is replaced by a substitute with a shift of specified positions down or up in the alphabet. Figure 2.1 shows an example of the shift cipher. The method is also known as *Caesar Cipher*.

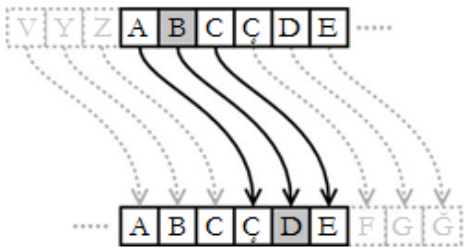

Figure 2.1 The Encryption Schema of the Shift Cipher

If each character of the alphabet is presented by an integer that corresponds to its position in the alphabet, the formula of the method for replacing each character *p* of the plaintext with a character *C* of the ciphertext with a shift of *k* in an alphabet consisting of *n* characters could be expressed as

$$C_i = (p_i + k) \bmod n \quad (2.1)$$

$$p_i = (C_i - k) \bmod n \quad (2.2)$$

As an illustrative example, the following text is chosen:

Plaintext: GELECEĞİKESTİREBİLMENİNENGÜZELYOLUONUİCATETMEKTİR

And by using the plaintext, it is encrypted as:

Ciphertext: IĞOĞEĞİLNĞUVLTĞDLOÖĞPLPĞPIZCĞOBROYRPYLEÇVĞVÖĞNVLT

The method is weak against to brute force attack. In worst case, an attacker could detect the plaintext in (n-1) steps.

2.1.2. Substitution Cipher

Substitution cipher depends on the principle that replacing each letter by another letter. Substitution characters are composed by a random permutation of the letters of the source language. *Caesar cipher* is a special form of a substitution cipher.

Table 2.1 Substitution Cipher Encryption Table

a	b	c	ç	d	e	f	g	ğ	h	ı	i	j	k	l	m	n	o	ö	p	r	s	ş	t	u	ü	v	y	z
E	R	T	Y	U	I	O	P	Ğ	Ü	A	S	D	F	G	H	J	K	L	Ş	İ	Z	C	V	B	N	M	Ö	Ç

Table 2.1 demonstrates an instance of substitution cipher. Each plaintext letter is looked up in the encryption table and written the corresponding ciphertext below. The decryption process depends on finding each ciphertext letter in the second row and writing down the corresponding letter from the top row.

As an illustrative example, the following text is chosen:

Plaintext : TEKNOLOJİSENDOĞDUKTANSONRAİCATEDİLENHERŞEYDİR

And by using the plaintext, it is encrypted as:

Ciphertext : VIJKGKDSZIJUKĞUBFVEJZKJİESTEVIUSGIJÜİİCIÖUSİ

The key space of substitution cipher is equal to n! on an alphabet consisting of n letters. For Turkish, the key space is equal to 29!, that is a number larger than 10^{30}. This is larger than many times of key space of *DES* which is equal to 10^{16}.

Suppose that the attacker is able to check a possibility per microsecond, it would take time more than 10^{16} years to solve ciphertext in the worst case[1]. Obviously, the encipherment rules out a brute force attack.

However, the attacker does not need to check all possibilities to solve ciphertext. It is a fact that the characteristic distribution of the texts is similar. Table 2.2 and Figure 2.2 illustrate the unigram frequencies of Turkish language. Therefore attacking by using an analysis of frequencies can lead to reveal encryption table.

Table 2.2 Percentage of Frequencies of Turkish Unigrams [2]

Letter	**Freq.**	**Letter**	**Freq.**	**Letter**	**Freq.**
A	11,92	I	5,114	R	6,722
B	2,844	İ	8,6	S	3,014
C	0,963	J	0,034	Ş	1,78
Ç	1,156	K	4,683	T	3,314
D	4,706	L	5,922	U	3,235
E	8,912	M	3,752	Ü	1,854
F	0,461	N	7,484	V	0,959
G	1,253	O	2,476	Y	3,336
Ğ	1,125	Ö	0,777	Z	1,5
H	1,212	P	0,886		

[1] $\frac{10^{30}10^{-6}}{60.60.24.265} > 10^{16}$

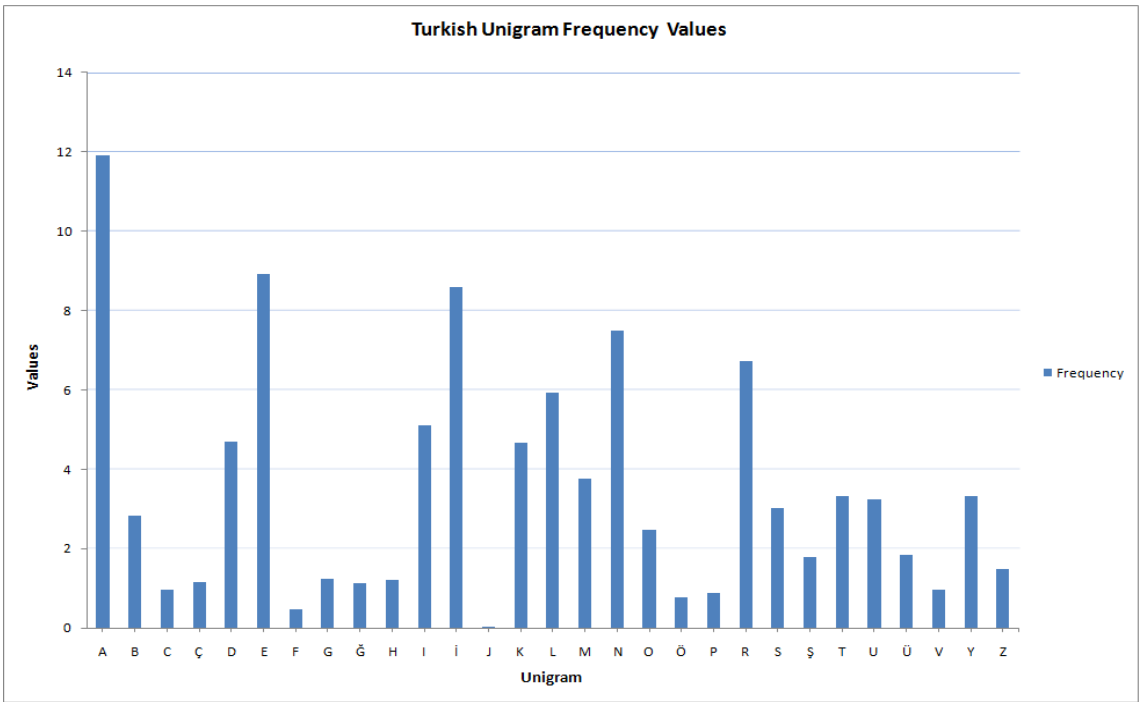

Figure 2.2 An Illustration of Frequency Values of the Turkish Unigrams

However, the letter frequencies of the ciphertext could not exactly match the letter frequencies of the source language in a short message. Looking at the most common words as listed in Table 2.3 and the most common n-grams is a valuable tool to encrypt the data. The rest of the operation depends on the heuristic approach. The structure of the words and stream of the sentence would help to detect the plaintext.

Table 2.3 The Most Frequent 100 Turkish Words [1]

Word	Word	Word	Word
Bir	Mi	Önce	Biz
Ve	İki	Nın	Vardı
Bu	Değil	İyi	Oldu
De	Gün	Onu	Aynı
Da	Büyük	Doğru	Adam
Ne	Böyle	Benim	Ancak
O	Nin	Öyle	Olur
Gibi	Mı	Beni	Ona
İçin	In	Hem	Biraz
Çok	Zaman	Hemen	Tek
Sonra	İn	Yeni	Bey
Daha	İçinde	Fakat	Eski
Ki	Olan	Bizim	Yıl
Kadar	Bile	Küçük	Bunu
Ben	Olarak	Artık	Tam
Her	Şimdi	İlk	İnsan
Diye	Kendi	Olduğunu	Göre
Dedi	Bütün	Şu	Uzun
Ama	Yok	Kadın	İse
Hiç	Nasıl	Karşı	Güzel
Ya	Şey	Türk	Yine
İle	Sen	Olduğu	Kız
En	Başka	İşte	Biri
Var	Onun	Çocuk	Çünkü
Türkiye	Bana	Son	Gece

Table 2.4, Table 2.5, Table 2.6 and Table 2.7 illustrate the most common n-grams in Turkish. The n-gram data presented below is collected from the data set of size 13.4 MB.

Table 2.4 Frequencies of 600 Most Frequent Bigrams in Turkish within 11M

Ngr	Fre	Ngr	Fre	Ngr	Fre	Ngr	Fre	Ngr	Fre	Ngr	Fre
AR	210907	SA	53696	İÇ	31429	Hİ	18821	NR	10949	EP	6092
LA	193023	RE	53009	RK	29528	ĞU	18753	İP	10803	KÖ	5975
AN	190681	Mİ	52476	UL	29371	ŞL	18154	OĞ	10648	KK	5957
ER	167490	DU	52283	KT	28964	ÇI	18112	IS	10610	BI	5760
İN	166362	TI	51386	RU	28453	ÖR	17778	GA	10575	NG	5692
LE	147573	Tİ	50968	SO	27989	ÇE	17736	Vİ	10524	EB	5667
EN	145677	AT	50661	AB	27564	ÜZ	17636	CI	10287	YR	5622
DE	138507	SI	50362	IY	27516	Şİ	17282	SÖ	10196	İF	5528
IN	132807	ET	49927	OK	27389	ÇA	17193	PE	10171	ÜÇ	5469
DA	127637	VE	49863	Çİ	27377	UĞ	16907	ÜS	10146	OM	5460
Bİ	125824	ED	49401	LU	27338	ZL	16237	US	10074	YN	5457
İR	120936	NL	47564	AP	27085	İB	15851	IP	10041	ÇL	5262
KA	113213	AS	47413	ÜR	27057	RS	15816	NS	9885	ZU	5205
YA	107833	AŞ	46294	ŞT	26972	ÖN	15278	EZ	9758	AA	5193
Dİ	101972	ON	45874	AĞ	26617	UZ	15172	LK	9755	RŞ	5193
MA	100389	BE	45640	UM	26139	UK	15139	CU	9701	ÇÜ	5186
ND	98136	KE	45296	KO	25764	ÇO	14880	ŞM	9568	RB	5141
RA	94205	SE	44712	EV	25568	ÖY	14744	TO	9312	OŞ	5093
AL	92166	BU	44624	DÜ	25363	UŞ	14641	BO	9184	OP	5075
AK	87361	EY	44573	GÖ	25203	ZI	14561	NT	8977	PT	5061
Rİ	84497	IR	43789	YI	24913	YÜ	14156	EÇ	8691	RO	4899
İL	76889	KL	43444	IĞ	24106	Zİ	13862	PI	8685	MS	4790
Nİ	73277	ÜN	43059	RM	24077	RÜ	13654	OY	8640	OS	4774
BA	71504	EM	42649	ZA	23774	TM	13595	FE	8636	PO	4765
RD	71256	IK	40865	TÜ	23620	EŞ	13465	ZD	8012	VL	4745
AY	70636	GE	40841	KU	23617	AV	13357	EF	7925	UP	4701
OR	70300	IL	39898	VA	23616	ÜK	13284	ÜT	7852	BÖ	4683
NI	69437	ES	39387	MU	23551	LÜ	13272	NK	7820	Aİ	4625
Lİ	69339	İK	38939	IZ	23091	YD	13222	SL	7774	OD	4565
ME	68578	İŞ	38227	ŞI	22692	UT	13198	LT	7738	ĞÜ	4516
RI	66743	UR	38171	ZE	22563	ÜL	13181	SÜ	7729	NB	4482
TA	65232	ĞI	37718	TU	22533	NM	13100	KS	7683	ÜĞ	4463
NE	64209	LD	37375	ŞA	22239	ÜŞ	13030	OT	7353	NN	4404
EL	64209	İS	37263	LL	22196	AÇ	12668	ĞR	7347	GU	4374
AM	63808	IM	37245	SU	22168	ÖZ	12477	OC	7324	TR	4288
EK	62698	CE	37156	İĞ	21975	DO	12350	EH	7137	SS	4284
DI	62359	RL	36959	NC	21614	ML	12350	KM	7097	LS	4266
YO	61044	MI	36784	PA	21525	İD	11953	PL	7068	LO	4242
Kİ	59905	NU	36391	HE	21182	FA	11877	ĞL	7029	RÇ	4136
UN	59182	Ğİ	36283	İT	20510	ŞK	11863	ZÜ	7025	ID	4104
İM	59020	Gİ	35787	UY	20367	MÜ	11630	DÖ	6918	FI	4036
AD	58667	İZ	35238	RT	20288	ÜM	11563	ÖL	6801	SM	4030
İY	57222	CA	33993	GÜ	20124	Cİ	11387	İH	6744	UD	3968
HA	56075	AH	33171	Yİ	19854	BÜ	11348	ĞA	6727	İC	3936
Sİ	55239	AZ	32755	ŞE	19742	MD	11337	SK	6723	VR	3925
YE	55184	IŞ	32622	TL	19492	KÜ	11279	ŞÜ	6549	NÇ	3916
NA	54770	YL	32373	EC	19480	ÜY	11253	Fİ	6520	VU	3907
OL	54389	LM	32197	EĞ	19472	YU	11219	LG	6372	Pİ	3837
TE	53881	ST	31918	TT	19323	AF	11150	RG	6210	HI	3786
LI	53814	KI	31841	AC	18985	NÜ	10971	ŞU	6102	KR	3777

NY	3776	ZG	2195	IV	1134	CO	672	NŞ	413	AO	234
HU	3772	TS	2172	GR	1123	EE	655	VS	401	GL	231
ĞE	3688	OZ	2171	MM	1121	ŞO	644	Uİ	393	BN	227
HT	3651	HR	2058	HS	1106	YH	640	OÇ	393	TN	223
ÖĞ	3603	LY	2043	MR	1098	VI	639	YC	391	KG	222
KÇ	3575	KŞ	2033	ZS	1088	UF	629	ZK	382	TD	220
RC	3536	UB	2008	SP	1075	VM	602	ÇS	381	PY	218
HO	3481	ĞM	1998	ÜF	1070	TB	596	ĞC	381	SF	208
ÖT	3447	PS	1994	ŞS	1066	ŞV	596	ÇR	372	CC	203
ZM	3369	NZ	1976	JA	1054	OG	593	TV	365	JD	201
UH	3327	HÇ	1975	IÇ	1042	OH	571	PÇ	362	IO	200
CÜ	3312	ZC	1923	HK	1029	VC	568	EJ	361	VH	200
YM	3282	FR	1908	ŞB	1028	ÇK	558	HZ	357	ŞN	197
UC	3174	ÖK	1877	VK	1021	FF	536	FS	352	LF	195
HM	3131	RP	1853	RY	1008	MŞ	534	ŞR	349	VZ	183
ÇT	3122	LB	1820	RV	1005	MZ	533	Oİ	344	NH	183
ÇM	3122	PK	1805	RF	991	LV	533	VO	332	HY	178
UÇ	3043	MC	1777	ÇÖ	990	HŞ	529	KB	330	LZ	177
IC	2972	OB	1620	UG	963	LP	523	YY	327	ÇG	176
ÜC	2954	DR	1594	YF	960	IA	522	MG	327	BZ	170
PR	2926	TÇ	1573	ÖF	954	CH	519	FY	326	IG	166
İV	2888	ÖM	1570	ÖB	938	ÖD	516	ĞN	324	KF	165
IT	2823	LC	1541	MP	931	AE	516	ŞG	321	JL	163
YG	2798	YK	1531	DL	916	EG	514	VŞ	319	OU	163
DD	2723	VG	1525	OV	906	KN	514	HD	318	CK	161
FO	2692	BR	1522	YV	896	Eİ	513	LH	317	JU	156
HL	2646	HB	1515	Jİ	883	CR	502	CD	315	ZT	155
PM	2626	TK	1512	ŞY	876	AU	495	TY	315	DN	154
VD	2621	OF	1492	ĞD	864	SN	491	UE	301	NV	152
FT	2587	ÜV	1476	YT	859	FÖ	483	BB	300	İJ	151
MO	2566	ÇU	1434	MH	851	SH	481	ZR	298	SC	143
GI	2534	YB	1426	RR	842	ÜB	481	ÜH	297	CB	142
LN	2528	ÖV	1414	MK	815	IH	478	SB	292	ÖH	137
ÜD	2508	PU	1407	RZ	814	IB	478	ÖC	291	UO	134
MB	2470	ÖŞ	1381	VV	808	KY	473	KH	290	ÖÇ	132
ÜP	2425	UA	1381	TÖ	804	ZZ	468	VN	287	TG	131
ÇB	2412	İG	1369	ĞZ	799	EO	467	İO	284	ŞŞ	129
FL	2401	LÇ	1352	VÜ	791	FÇ	456	YZ	279	ÇH	126
ZO	2386	ZY	1351	PÜ	784	EA	448	ŞF	277	HH	126
NO	2385	AJ	1341	AG	745	HP	447	VY	274	TP	125
ÖP	2371	YÖ	1319	GO	739	İE	446	NP	274	OE	118
ÖS	2368	BD	1292	TF	730	OJ	438	ZN	273	İU	118
RH	2368	HV	1289	ZB	726	RÖ	435	ĞS	270	PP	118
SY	2367	FU	1275	PH	713	MY	434	ÜE	270	JO	117
RN	2353	FK	1263	OO	713	KD	432	NF	262	ĞB	114
UV	2325	SR	1257	BL	710	ŞH	425	RJ	250	KV	114
HÜ	2320	ŞÖ	1235	TH	709	SV	425	II	250	TC	109
İİ	2309	İA	1235	HN	709	MT	423	NÖ	249	MF	108
YS	2280	MN	1199	ŞÇ	703	CL	416	UU	235	YŞ	102
FÜ	2208	IF	1137	DY	683	JE	415	JI	235	ZH	97

Table 2.5 Frequencies of 600 Most Frequent Trigrams in Turkish within 11M

Ngr	Fre	Ngr	Fre	Ngr	Fre	Ngr	Fre	Ngr	Fre	Ngr	Fre
LAR	89715	ORD	18348	SAN	12433	ÇOK	9685	AŞK	7880	KTI	6634
BİR	77192	İYE	18295	YAP	12361	CAK	9670	ĞUN	7798	ĞIM	6633
LER	68463	BAŞ	18165	ORU	12334	RDA	9602	DIN	7781	İRD	6624
ERİ	58199	EDE	18132	AŞI	12303	DİM	9595	İKİ	7729	YET	6601
ARI	56640	ALI	17950	AHA	12211	EKL	9211	YEN	7718	ÜRÜ	6600
YOR	46374	UNU	17894	GÖR	12199	AKI	9195	LDİ	7710	LAD	6597
ARA	39442	END	17848	EMİ	12047	MAS	9191	ŞTİ	7687	RAD	6581
NDA	38533	BEN	17793	DİĞ	11989	ARK	9173	TEN	7679	BİZ	6558
İNİ	36926	NDİ	17595	VER	11988	OLM	9072	KTA	7655	ŞKA	6553
INI	34233	ÇİN	17466	RDİ	11935	HER	9055	RMA	7648	ÜND	6552
RİN	32092	ELİ	17380	RKE	11897	LDI	9052	OLU	7646	DİN	6550
DEN	31700	MIŞ	16453	OLD	11673	GÜN	8955	ENE	7566	OKU	6541
AMA	31046	EĞİ	16136	GİB	11642	AŞL	8928	DIR	7561	İRL	6541
NDE	30920	RLA	16134	AKL	11636	RUM	8917	SEN	7524	RAS	6510
EDİ	29711	MAN	16042	LEN	11505	URU	8887	IRA	7521	TİN	6489
ANI	29369	YAN	15605	KAN	11473	YLA	8818	İLM	7493	LİR	6475
ASI	28332	İLİ	15586	İMİ	11433	YER	8793	DAK	7486	ĞİL	6469
DAN	28109	SON	15500	MAK	11343	MAD	8788	ETT	7422	KTE	6450
NLA	28015	IYO	15455	ÇIK	11312	DEĞ	8728	YIL	7367	EKT	6425
AYA	27995	ONU	15454	STE	11290	İŞT	8673	MEY	7323	RKA	6408
RIN	27197	ECE	15382	DER	11222	DAR	8657	AŞA	7319	ALD	6403
IND	26780	KAD	15378	DED	11197	PAR	8558	GEÇ	7184	IRI	6402
ENİ	26210	UĞU	15363	DIĞ	11092	ŞEY	8555	HAL	7150	CEK	6396
ADA	25992	EME	14961	MEK	11090	YAZ	8511	API	7113	DİK	6394
İLE	25278	İST	14871	ĞİN	11013	TÜR	8501	ISI	7107	ALİ	6378
İND	24708	AĞI	14823	NUN	10991	AZI	8491	YAT	7107	KUR	6374
ALA	24559	OLA	14792	ŞLA	10966	REK	8457	NLI	7106	ÜNÜ	6358
NIN	23573	INA	14787	RLE	10783	NLE	8440	ERK	7084	LIĞ	6354
ANL	23470	ERD	14758	AND	10735	AKA	8404	HİÇ	7076	MED	6340
KAR	23010	ANA	14707	İSİ	10718	ULA	8404	TAR	7068	TUR	6339
LAN	22870	AYI	14639	DUĞ	10666	IMI	8366	KON	7066	NDI	6336
IĞI	22842	MİŞ	14564	MAY	10661	DAH	8348	SİZ	7050	GER	6326
ADI	22334	KAL	14496	RDE	10552	LAY	8336	RME	7010	BUN	6301
SIN	22148	LMA	14485	NRA	10535	TIR	8325	YAL	6996	ĞİM	6279
SİN	21714	DİY	14100	ONR	10532	HAY	8312	LİY	6993	DİL	6269
ESİ	21362	EYE	13914	LAM	10391	ARL	8309	BAN	6993	ALL	6264
YLE	21319	ÖYL	13810	NCE	10390	STA	8261	LİN	6979	APA	6244
KEN	21223	İRİ	13799	LİK	10323	ULU	8249	UND	6965	LEM	6223
RDU	21110	RDI	13746	LLA	10321	AKT	8218	ART	6934	DİR	6200
ELE	20913	RAK	13357	AMI	10253	ERL	8198	EVİ	6931	KAY	6196
İYO	20215	KLE	13276	VAR	10250	YAR	8153	DÜŞ	6913	HAN	6158
İĞİ	20109	İBİ	13215	EYİ	10232	IŞT	8101	İŞİ	6894	LAT	6084
İNE	19994	ABA	13100	ŞTI	10166	RAN	8039	DIM	6881	İYİ	6065
ARD	19356	ATI	12815	BUL	10087	TİR	7980	TLE	6795	IŞI	6043
KLA	19224	LDU	12749	AKİ	10042	TER	7975	TÜN	6788	MİN	6041
İÇİ	19160	ILA	12665	LIK	10011	İZİ	7951	ATL	6759	KİM	6030
TAN	18914	EKİ	12626	BAK	10007	ÜZE	7943	DUR	6718	KAP	6015
NİN	18606	ACA	12607	ETİ	9978	REN	7922	ZAM	6688	URA	5976
BİL	18502	ĞIN	12555	SUN	9885	IKL	7896	MUŞ	6665	YAK	5970
ERE	18440	GEL	12534	TTİ	9836	ILI	7881	SÖY	6645	NİZ	5970

RAL	5950	İRE	5347	LİM	4722	GİR	4301	OCU	3910	İTA	3599
MİZ	5940	İLD	5343	IZI	4709	ÜYO	4299	EKE	3906	LTI	3599
SEV	5934	MIZ	5288	NER	4696	RAM	4292	ÜRE	3906	SAY	3588
TIN	5920	ABİ	5264	ILD	4693	RIL	4243	TİK	3904	KKA	3587
İKA	5915	IKA	5253	KİL	4674	EVE	4233	URD	3904	TEM	3583
ATA	5895	MET	5247	CAĞ	4672	LAC	4210	TİY	3902	NNE	3583
UYO	5892	ORL	5241	TAK	4655	SAL	4207	ÖRE	3901	LLİ	3581
İRA	5890	AVA	5235	IRL	4650	YDI	4205	RŞI	3901	BÜY	3577
ÜRK	5850	RES	5233	ACI	4645	LİĞ	4201	IKI	3897	NIM	3571
ASA	5849	KAT	5228	MAZ	4629	SIL	4194	TMA	3892	SEL	3567
İDE	5841	ETE	5182	ÖRÜ	4600	MIN	4180	BER	3883	EBİ	3563
YOK	5827	RİY	5156	ALM	4600	DUY	4180	RİL	3878	TTI	3547
TME	5817	ANM	5153	IYL	4589	HEM	4177	KIR	3874	TAL	3546
ETM	5816	MER	5148	ALT	4586	YİN	4163	ÜZÜ	3851	MEL	3545
BEL	5815	AZA	5111	İNS	4586	IKT	4160	ÇOC	3833	ÖNÜ	3537
TEK	5813	LIN	5066	DÖN	4583	ÜĞÜ	4160	RAY	3833	ORA	3529
NAN	5808	ÜST	5059	MİY	4573	MUT	4153	YDİ	3832	ONA	3524
LME	5807	SOR	5046	ZER	4571	LMİ	4144	NME	3829	BAH	3523
MES	5803	MAM	5039	ŞAR	4566	TIL	4134	NİY	3819	DÜR	3520
TLA	5792	İYA	4988	RET	4558	EVL	4123	ADE	3813	OYU	3519
GEN	5791	LEY	4987	ELL	4554	ARE	4118	LEC	3809	MDE	3517
ÜŞÜ	5781	UKL	4931	İYL	4550	GİD	4114	ANK	3808	INL	3494
ZLE	5781	LED	4930	UMU	4525	ŞİM	4112	ŞTU	3803	AYR	3493
ŞIN	5779	KLI	4929	NSA	4519	STİ	4110	SIZ	3789	AYN	3478
RİM	5772	RAR	4927	ŞÜN	4498	ÜYÜ	4106	HAR	3783	RMİ	3475
MAL	5770	LIY	4922	İZL	4488	KLİ	4104	UZU	3780	İLG	3468
LAŞ	5765	KIZ	4908	TİM	4477	APT	4089	KTİ	3770	MEZ	3467
GÖZ	5755	MEM	4893	KOR	4476	TIM	4085	ERA	3754	BEK	3453
ÇEK	5729	RTI	4887	ŞLE	4460	ENL	4084	YOL	3749	RIY	3439
KIN	5699	RLİ	4887	DOĞ	4452	KİN	4080	İŞL	3743	LİS	3437
UNA	5677	DUM	4886	LAH	4450	UNL	4059	IRM	3742	DOL	3435
ATT	5657	İNC	4866	NEM	4436	KER	4042	ARM	3735	ZAR	3426
MEN	5641	CAN	4864	ZLA	4434	BAB	4017	NED	3722	ANB	3423
ŞMA	5637	BEY	4863	İRM	4431	ÖNC	4016	LMI	3721	DAY	3420
ÜTÜ	5629	ILM	4860	KES	4418	AĞA	4008	KAS	3718	İML	3408
İKL	5622	PLA	4859	CEĞ	4410	RDÜ	3998	ANN	3715	NEL	3406
TIK	5614	TIĞ	4856	SES	4392	BİN	3983	EŞİ	3699	İZE	3395
TEL	5589	BÜT	4849	İNL	4392	BAL	3973	ELD	3697	OTU	3383
MLE	5585	ÇAL	4831	GİT	4392	MLA	3966	SİY	3688	TAB	3381
NIZ	5575	ARİ	4831	KUL	4389	SAR	3960	IZL	3682	AST	3370
NCA	5570	BAS	4829	İKT	4375	SÜR	3954	PEK	3670	İDİ	3353
YÜZ	5561	DAM	4827	YÜK	4374	IRD	3932	SER	3661	LUK	3349
USU	5550	ZEL	4823	UYU	4351	KAÇ	3930	İNA	3654	SEY	3345
DEM	5541	ONL	4821	NMA	4344	MDA	3927	NAS	3646	OCA	3344
İMD	5524	NİM	4793	ÜNE	4338	ERM	3920	ZİN	3644	RAF	3343
LLE	5489	LIŞ	4790	ANT	4327	ANE	3918	YON	3623	ORT	3330
YAŞ	5446	ERS	4786	TİĞ	4325	DIK	3917	SIR	3620	KET	3327
HAT	5435	AÇI	4762	ARŞ	4323	ÖNE	3917	UŞT	3615	PTI	3326
DEK	5412	LUN	4753	AYD	4317	EYL	3914	LET	3615	EFE	3324
RIM	5358	ETL	4733	RTA	4302	BUR	3912	ALE	3608	RSU	3316

Table 2.6 Frequencies of 600 Most Frequent Tetragrams in Turkish within 11M

Ngr	Fre	Ngr	Fre	Ngr	Fre	Ngr	Fre	Ngr	Fre	Ngr	Fre
LARI	43055	LARA	7760	BİLE	5796	İYLE	4546	LİĞİ	3804	GÖRE	3399
LERİ	36077	BENİ	7737	RİNE	5782	IĞIM	4538	IRLA	3775	KAPI	3395
ERİN	26911	YORU	7724	EREK	5771	ÜŞÜN	4495	EDİM	3763	İNİZ	3380
ARIN	26117	ILAR	7684	LIĞI	5697	İĞİM	4494	ILDI	3759	ŞLER	3377
INDA	25969	AMAN	7633	ŞLAR	5661	ENİM	4474	ASIL	3731	YAZI	3376
İNDE	22658	MAYA	7631	MIŞT	5623	LEDİ	4378	İRİN	3728	ANIM	3367
İYOR	19578	EDEN	7603	ONUN	5615	SİNE	4361	ABİL	3726	OLAR	3366
ORDU	17770	İĞİN	7583	ECEK	5589	CEĞİ	4351	LECE	3715	İNLE	3361
ANLA	17328	AŞLA	7523	NLER	5532	ECEĞ	4324	SINA	3715	RDUM	3356
İÇİN	16930	ARDA	7507	ANDI	5526	KALA	4312	AYAT	3715	LANI	3327
NLAR	16745	SINI	7449	MADI	5480	KARI	4301	ETLE	3715	ALTI	3319
YORD	16018	ARAK	7434	ĞUNU	5384	UNUN	4298	IMIZ	3694	URDU	3315
ENDİ	15616	ERDE	7387	İSİN	5370	TİĞİ	4288	NLAT	3693	NASI	3286
IYOR	15398	KARA	7375	LAMA	5364	ÜYOR	4287	LERE	3686	SENİ	3286
ASIN	14773	SIND	7361	ADAN	5327	RDEN	4284	STER	3685	MİYO	3281
ÖYLE	12783	OLMA	7267	ALDI	5282	MESİ	4237	LAYA	3680	TANB	3251
KLAR	12718	UĞUN	7229	TÜRK	5234	ACAĞ	4232	ARTI	3678	ANBU	3240
ALAR	12303	ACAK	7067	RIND	5201	YENİ	4232	ISIN	3676	NBUL	3236
NDAN	12068	ERDİ	6895	MEYE	5166	ERİM	4226	ARAS	3674	IRDI	3222
ANIN	12011	RİND	6817	HAYA	5148	ARIM	4211	ULAR	3658	İSTİ	3212
GİBİ	11641	ENİN	6815	RADA	5137	ÇİND	4209	TANI	3654	LEME	3205
DİĞİ	11316	ERLE	6737	KONU	5135	AŞIN	4204	VARD	3642	GELE	3203
OLDU	11130	İSTE	6723	BAŞL	5124	MEDİ	4189	KARŞ	3635	AKLI	3202
DIĞI	11066	ĞINI	6721	ADIĞ	5096	SİND	4154	ADAM	3615	ETİN	3186
ININ	10737	SÖYL	6638	RASI	5053	ETME	4153	IKAR	3612	APTI	3185
DUĞU	10666	ORUM	6634	ÜTÜN	5028	LLAR	4143	RDAN	3608	EYLE	3174
İLER	10665	RKEN	6627	DEKİ	5003	ZLER	4135	ÇALI	3591	ATIR	3167
ESİN	10658	DAKİ	6596	MİŞT	4976	ALAN	4092	LIYO	3585	İMDİ	3141
ELER	10580	BİRİ	6585	ARKA	4914	AĞIN	4088	BÜYÜ	3575	TELE	3134
DEDİ	10528	ERKE	6575	ORLA	4901	YORL	4081	İLME	3568	ENLE	3130
ONRA	10527	MASI	6487	EDİĞ	4845	KTAN	4080	ALLA	3567	KORK	3116
SONR	10525	EĞİL	6466	İMİZ	4816	GÖRÜ	4073	AKTA	3559	İRLE	3107
ARDI	10515	İŞTİ	6459	TIĞI	4806	AYAN	4065	İKTE	3553	YAPA	3098
İNİN	10499	LERD	6405	İKLE	4799	ALIŞ	4025	ELİN	3547	ANNE	3096
KEND	10462	RLER	6398	BÜTÜ	4768	İNSA	4014	EMİŞ	3540	BİLM	3092
NDEN	10226	IKLA	6377	ÜNDE	4766	AMIŞ	4013	UNLA	3513	GENE	3092
RİNİ	10148	ĞİNİ	6342	İRDİ	4765	NSAN	4000	LMIŞ	3509	ELDİ	3088
LARD	10110	LADI	6309	DÜŞÜ	4740	BANA	3998	DİLE	3507	OCUK	3087
KADA	9380	ARLA	6278	LİYO	4698	BAŞI	3993	BAKA	3505	ATTI	3086
RINI	9329	ZAMA	6262	DİYO	4691	İSTA	3990	LACA	3488	SUNU	3086
RLAR	8666	EKLE	6261	ONLA	4676	AKTI	3981	YAPI	3483	KANI	3081
KLER	8628	BAŞK	6236	ÇIKA	4635	OLAN	3964	EDER	3476	AĞIR	3075
DİYE	8587	ETTİ	6220	YANI	4629	ÖNCE	3964	EMEK	3465	İLMİ	3066
SİNİ	8576	ADAR	6171	KADI	4625	BABA	3933	İLİR	3464	EVLE	3053
IĞIN	8441	AŞKA	6152	CAĞI	4623	ARŞI	3901	UŞTU	3451	İLGİ	3052
LDUĞ	8212	ANLI	6125	STAN	4608	BİRL	3843	NDAK	3434	ÜZEL	3045
DAHA	8071	ADIN	6108	İLDİ	4583	ÇOCU	3833	İRLİ	3421	RESİ	3045
AKLA	7957	BİLİ	6067	TLER	4583	BULU	3821	EĞİN	3420	RINA	3042
IŞTI	7948	UYOR	5887	IYLA	4575	GELİ	3817	ÜZER	3420	DENİ	3038
DEĞİ	7914	UNDA	5828	ANDA	4574	UKLA	3813	İŞLE	3412	ALMA	3029

ANMA	3024	AYIN	2732	OYUN	2477	MIYO	2304	TARA	2190	ERÇE	2063
BÖYL	3012	KİTA	2727	USUN	2473	YECE	2300	LTIN	2185	HİSS	2062
GECE	3002	GELD	2722	ULLA	2471	ÜSTÜ	2298	ĞIND	2182	MAKT	2062
EDİY	2997	DÜĞÜ	2719	İZLİ	2470	LMAD	2295	NDİN	2181	YLED	2054
YORS	2984	ULUN	2708	EYİN	2461	EKTE	2290	HATI	2176	LUĞU	2046
YAŞA	2983	DURU	2702	NİYE	2461	LMAK	2289	YLER	2176	ANCA	2044
VERİ	2978	LİKT	2695	EMİN	2459	LDİĞ	2287	IKTI	2175	LMAY	2043
İTTİ	2978	IRAK	2688	ALNI	2459	ÖSTE	2286	NESİ	2173	İSSE	2041
NDİM	2978	ILMA	2682	ORSU	2453	LNIZ	2286	ILMI	2171	NIYO	2037
TLAR	2974	LIKL	2680	İDEN	2452	YALN	2286	KALI	2170	FEND	2033
DİSİ	2971	EKLİ	2672	SANI	2448	KİŞİ	2283	İNAN	2164	MELE	2026
YAPT	2956	AKIN	2666	ÜZÜN	2446	SONU	2283	AKAL	2164	RABA	2026
OTUR	2945	MADA	2663	MALA	2446	RİMİ	2281	MAMI	2164	TMİŞ	2025
ŞİMD	2940	LMAS	2652	ÖĞRE	2436	ÇBİR	2276	RDIM	2155	GELM	2025
LLER	2925	ERİY	2652	EYEC	2427	RECE	2276	İNCİ	2155	EFEN	2025
BİZİ	2919	RTIK	2648	DERİ	2422	ŞIND	2275	YERE	2153	SÜRE	2024
NEDE	2908	EVİN	2646	AYAC	2421	GÖZL	2267	ZETE	2147	EDİK	2024
NDEK	2906	ELİM	2641	AYLA	2415	ANIY	2266	RAYA	2146	OLUR	2023
ONUŞ	2903	LANM	2641	RMİŞ	2413	AMAY	2263	AKAN	2140	YÜRÜ	2023
LAMI	2902	NLIK	2638	GÖRD	2411	SORU	2260	SIRA	2138	TILA	2022
ÜYÜK	2902	ATIN	2636	MLAR	2410	LDIR	2260	İMSE	2135	YAKI	2022
ADIM	2887	ELLİ	2624	BEKL	2408	MAYI	2258	İŞTE	2131	SEVİ	2021
NDİS	2884	EKTİ	2619	UZUN	2406	YAPM	2258	ŞKAN	2126	ORKU	2019
ORTA	2881	LMİŞ	2611	HMET	2403	AMAD	2257	GETİ	2126	SOKA	2016
ZERİ	2875	ARAR	2607	İLEN	2402	BURA	2257	AZET	2124	LANA	2015
YATI	2873	YÜZÜ	2595	KASI	2402	ENCE	2256	ARKE	2124	MASA	2014
YACA	2867	LEYE	2591	LARL	2394	ERME	2255	URMA	2120	ARAY	2009
TTİĞ	2865	AYDI	2585	İRME	2391	ILLA	2250	HEME	2119	İZLE	2006
MANI	2857	İMLE	2579	EBİL	2388	GİDE	2248	KALD	2118	SAAT	2005
ATLA	2853	ÖRDÜ	2573	EMEN	2374	GÖST	2246	ŞLAD	2116	ZLAR	2004
İYET	2824	İNCE	2562	ARAN	2373	ELİR	2246	GAZE	2115	RMIŞ	2000
LLAH	2822	DOĞR	2562	BIRA	2373	REDE	2244	IZLA	2113	ARAB	1994
ÖZLE	2819	APIL	2560	AMIN	2372	DEME	2239	İRAZ	2113	KARD	1993
KTEN	2817	YARA	2553	HİÇB	2367	ALIN	2239	MUTL	2104	YARI	1986
RALA	2805	GEÇİ	2546	ĞIMI	2366	IŞLA	2237	İYDİ	2100	GÖRM	1985
PARA	2802	ATLI	2543	İÇBİ	2365	LDIĞ	2232	LİKL	2099	ATTA	1983
RSUN	2800	DEMİ	2541	GERE	2358	ARMA	2231	KESİ	2098	ERİL	1979
MLER	2784	OĞRU	2540	YILL	2356	DİNİ	2224	ARAL	2095	IKTA	1979
İLİY	2783	ELEN	2531	İMDE	2347	İZİN	2222	AMLA	2088	KİYE	1977
BAKI	2781	EYEN	2512	ILAN	2345	KANL	2222	DINI	2087	NERE	1972
GÜZE	2774	INIZ	2507	IRMA	2341	YERİ	2222	ATIL	2087	BİRB	1969
BUNU	2773	LAND	2504	RDİĞ	2332	ALIK	2219	AĞIM	2087	AYRI	1968
RKAD	2769	UĞUM	2499	LMUŞ	2331	OLAY	2217	İZİM	2082	ÜNKÜ	1963
RLİK	2762	ARIY	2497	GİTT	2324	ETİR	2208	RMUŞ	2077	AZIL	1963
ESKİ	2751	ADAŞ	2497	KÜÇÜ	2320	LERL	2202	IMDA	2074	HANE	1962
MEKT	2750	LAYI	2493	ABAS	2316	ENİZ	2202	İLEC	2068	İRMİ	1961
ŞTIR	2749	VERD	2488	ÜNYA	2315	VERM	2199	ĞREN	2068	MIZI	1961
ŞTİR	2743	BASI	2484	AĞLA	2315	AŞTI	2198	ENDE	2067	RBİR	1956
EDİL	2737	INLA	2484	DÜNY	2311	TEDİ	2195	ERSE	2066	RIMI	1956
MANL	2735	ĞİMİ	2480	AYNI	2307	BİRA	2193	CUKL	2064	AZAN	1953

Table 2.7 Frequencies of 600 Most Frequent Pentagrams in Turkish within 11M

Ngr	Fre	Ngr	Fre	Ngr	Fre	Ngr	Fre	Ngr	Fre	Ngr	Fre
LARIN	23638	ONLAR	4418	BİLİR	2942	RLARD	2396	YILLA	2106	ALDIR	1841
LERİN	19659	ORLAR	4406	ASINA	2931	İRLİK	2386	GAZET	2105	ARIMI	1834
YORDU	16000	KADIN	4359	BÜYÜK	2902	LMASI	2377	BİLME	2103	KİMSE	1833
SONRA	10524	ECEĞİ	4269	KONUŞ	2897	BIRAK	2372	ŞLADI	2101	YÜZÜN	1827
KENDİ	10266	DÜŞÜN	4234	ORDUM	2889	LARLA	2371	KALDI	2081	AMAYA	1823
ARINI	9092	BENİM	4232	NDEKİ	2884	HİÇBİ	2364	AĞINI	2080	YAPIL	1818
KLARI	8788	ACAĞI	4212	DİĞİM	2866	İNDEK	2363	UYORD	2076	MAKTA	1817
ERİNİ	8428	İÇİND	4181	TTİĞİ	2862	LERİM	2357	ILMIŞ	2069	ADINI	1811
LDUĞU	8212	ÇİNDE	4147	LIĞIN	2857	İRLER	2352	BİLİY	2066	İĞİMİ	1806
INDAN	7993	YORLA	4081	ŞİMDİ	2839	RALAR	2337	AŞLAD	2063	STEDİ	1806
ANLAR	7940	SİNDE	4041	ENDİS	2830	RDİĞİ	2331	İLECE	2062	ANIND	1803
OLDUĞ	7706	İNSAN	3937	ERDEN	2823	GİTTİ	2323	ETTİĞ	2053	İSTED	1801
NLARI	7546	ESİNİ	3920	ZLERİ	2781	ÖZLER	2323	YLEDİ	2053	EĞİNİ	1801
SINDA	7329	LARDI	3858	NASIL	2776	LANDI	2317	LİKLE	2052	FÜSUN	1797
ALARI	7133	DİĞİN	3785	RKADA	2768	DÜNYA	2305	DİSİN	2039	İŞTİR	1796
İNDEN	7047	RASIN	3785	BİRİN	2768	IYORU	2301	HİSSE	2034	NDIĞI	1795
İYORD	6793	İSTAN	3711	ARKAD	2766	SININ	2300	ÖYLED	2025	LENDİ	1795
RİNDE	6778	HAYAT	3673	NDİSİ	2762	LIKLA	2299	LİĞİN	2025	AMIŞT	1793
DEĞİL	6254	KARŞI	3633	YANIN	2756	MIYOR	2299	BİZİM	2024	RLARI	1790
LARDA	6239	İYORU	3615	GÜZEL	2753	BULUN	2295	OCUKL	2023	ÇÜNKÜ	1786
ZAMAN	6158	ARASI	3592	DIĞIM	2742	LDİĞİ	2287	FENDİ	2023	KARAR	1785
ERİND	6144	LIYOR	3574	ENDİM	2727	YALNI	2286	NIYOR	2013	ANLAM	1780
KADAR	6113	DIĞIN	3560	GELDİ	2720	ALNIZ	2286	KORKU	2012	ÜNDEN	1768
BAŞKA	6106	ANLAT	3543	ÜZERİ	2716	EDİYO	2285	DEDİM	2002	AMADI	1768
YORUM	6049	VARDI	3535	UNLAR	2710	MESİN	2277	ALLAH	2002	IĞIMI	1763
ASIND	5970	İKLER	3516	OLARA	2693	İÇBİR	2272	ANLIK	2002	IŞTIR	1761
SÖYLE	5963	EKLER	3504	BİRLİ	2687	LACAK	2270	EFEND	1993	OLACA	1750
KLERİ	5814	ÇIKAR	3490	ETLER	2670	BAŞIN	2252	BİRAZ	1942	BAKAN	1743
ELERİ	5801	NDAKİ	3401	TLERİ	2669	GÖSTE	2246	NINDA	1933	İSTİY	1741
MIŞTI	5606	RİNİN	3366	ŞLARI	2653	ÖSTER	2246	BİRBİ	1927	MUTLU	1740
IYORD	5448	RININ	3337	İSİNİ	2645	MADAN	2232	İRBİR	1926	İLMİŞ	1736
DUĞUN	5346	MASIN	3326	İLİYO	2633	LDIĞI	2229	TİYOR	1920	AŞKAN	1734
UĞUNU	5192	ERLER	3323	LİKTE	2632	GEREK	2213	ÖĞREN	1915	ÇIKTI	1733
RINDA	5148	MİYOR	3267	MADIĞ	2607	ARLAR	2199	LLERİ	1912	ERKES	1733
ARIND	5101	ANBUL	3235	İŞLER	2599	ESİNE	2199	BUNLA	1912	HERKE	1726
ADIĞI	5089	STANB	3234	SİNİN	2587	ŞINDA	2188	GÖZLE	1904	İKAYE	1722
ERİNE	4906	TANBU	3234	ERKEN	2586	RLİKT	2188	BEKLE	1902	TÜRKİ	1719
LERDE	4894	ÇALIŞ	3219	INDAK	2579	NLERİ	2187	IŞLAR	1901	MUŞTU	1715
MİŞTİ	4892	ILARI	3176	DİLER	2556	ALTIN	2182	UNDAN	1891	RKİYE	1708
EDİĞİ	4844	LARAK	3175	ARTIK	2551	ILLAR	2180	LMADI	1881	YERİN	1707
AKLAR	4806	LARIM	3174	DOĞRU	2536	MALAR	2180	MANIN	1879	ÜRKİY	1706
BÜTÜN	4767	ARDAN	3095	ZERİN	2529	LERLE	2172	LAMIŞ	1877	ABASI	1706
LİYOR	4693	ÇOCUK	3085	KADAŞ	2485	ĞINDA	2161	KÜÇÜK	1876	MEDEN	1704
İLERİ	4689	RLERİ	3077	VERDİ	2475	AYATI	2149	TLARI	1865	LMAYA	1701
DİYOR	4671	ESİND	3039	ENLER	2450	ENDİN	2145	ATIRL	1863	BELKİ	1701
BAŞLA	4638	BÖYLE	3012	ORSUN	2448	HATIR	2145	SENİN	1854	ERİMİ	1699
IĞINI	4627	UKLAR	2999	YORSU	2438	MANLA	2129	CAĞIN	1853	TARİH	1696
ASINI	4582	ARINA	2989	NEDEN	2417	GETİR	2126	RİYOR	1846	İMLER	1695
İĞİNİ	4535	YAPTI	2948	AYACA	2417	HEMEN	2119	GERÇE	1846	LİNDE	1693
IKLAR	4519	AŞLAR	2947	GÖRDÜ	2408	AZETE	2106	ANIYO	1846	CEĞİN	1686

ĞİNDE	1679	LANMA	1540	ALMIŞ	1433	CUKLA	1311	ENDİL	1205	HAREK	1151
INLAR	1676	LECEĞ	1540	DIKLA	1423	ÇLARI	1310	ÜŞÜND	1202	NLATI	1151
ARKEN	1675	NLARD	1539	IKTAN	1422	SEYRE	1309	ZÜNDE	1200	OLMAK	1151
DUĞUM	1672	MEKTE	1539	PTIĞI	1418	SEVER	1309	ERMİŞ	1200	MİŞLE	1147
RANLI	1662	İSTEM	1538	GEÇİR	1411	TELEF	1309	LMIŞT	1200	RESİN	1145
EHMET	1658	KİTAP	1535	ÖYLEM	1410	ELEFO	1303	İRDİĞ	1199	LACAĞ	1144
STİYO	1658	GÖRME	1535	SANKİ	1410	LEFON	1303	KARIŞ	1199	YARDI	1143
MEHME	1655	OLMAD	1533	EMEDİ	1408	TOPLA	1301	LEDİĞ	1199	YORMU	1142
EYECE	1654	LINDA	1528	SONUN	1406	EKTEN	1299	DERDİ	1198	NDİMİ	1141
LLARI	1654	İNLER	1527	APTIĞ	1405	ARŞIL	1295	SİYLE	1198	ADAŞL	1138
MEDİĞ	1647	MELER	1527	BASIN	1404	YECEK	1291	TİĞİM	1193	ENCER	1137
NLIĞI	1644	ARIYL	1522	EMEYE	1399	ARMIŞ	1285	LADIM	1192	ŞÜNDÜ	1137
TINDA	1644	DEĞİŞ	1518	YERLE	1398	NDİNİ	1281	BİRDE	1190	NCERE	1134
YAPMA	1643	ANLIĞ	1516	MAYAC	1397	ERİYL	1280	AÇLAR	1190	NİDEN	1134
BİLDİ	1642	EMİŞT	1515	ANCAK	1395	YACAĞ	1280	ARADA	1190	AKLAŞ	1133
BİLEC	1639	CAĞIM	1511	ARALA	1388	ORADA	1280	LAMAY	1189	ELİYO	1133
ARABA	1636	İLGİL	1511	ARANL	1384	YATIN	1277	LEŞTİ	1189	ERDİĞ	1132
ERÇEK	1631	RİYLE	1511	İĞİND	1378	ALARD	1276	BAĞIR	1188	LANLA	1131
ANDIĞ	1631	LUYOR	1511	RİMİZ	1372	CAKLA	1275	AMIYO	1183	LÜYOR	1131
İYORL	1631	TİĞİN	1508	ETMİŞ	1370	ĞIMIZ	1273	BELİR	1182	LMESİ	1130
DİKLE	1630	LECEK	1507	ISINI	1369	ÜŞÜNÜ	1264	SORDU	1182	YÜKSE	1129
ANMIŞ	1626	ŞIYOR	1506	BABAM	1368	UŞTUR	1263	ADINL	1182	FAZLA	1129
ÜYORD	1622	LERDİ	1505	TARAF	1364	RDÜĞÜ	1261	IRLAR	1182	BENZE	1128
AŞIND	1620	RUYOR	1491	ÖYLEY	1363	TIĞIM	1259	KİLER	1181	RÜYOR	1128
OLMUŞ	1619	ÖLDÜR	1491	DUYGU	1360	İSTER	1258	BURAD	1179	YERDE	1126
DENİZ	1617	ŞLERİ	1489	RIYOR	1356	AZILA	1257	ONUŞM	1179	ALIŞI	1126
ULARI	1616	KASIN	1484	OLMAS	1355	URADA	1257	SESSİ	1178	GÖTÜR	1125
NSANL	1613	LAŞTI	1481	VERME	1353	ELLER	1247	ACAKT	1177	DEVLE	1125
AKTAN	1611	İYORS	1480	ERKEK	1352	RESİM	1246	ALDIĞ	1175	GEÇTİ	1125
ANDIR	1595	LABİL	1480	AMLAR	1348	YOKTU	1241	EDİLE	1174	OLSUN	1122
GELEN	1595	SOKAK	1479	İRDEN	1346	SİNİZ	1239	ESSİZ	1174	NİYET	1121
GÖRÜN	1592	BAKTI	1478	ISIND	1340	EMİYO	1238	URUYO	1169	ANDAN	1120
ŞEYLE	1590	KIYOR	1475	MLARI	1339	SUNUZ	1236	İKTEN	1169	GENEL	1119
TIRLA	1590	İMİZİ	1474	TİRDİ	1339	LAYAN	1236	IMIZI	1168	CİLER	1119
ÜSTÜN	1587	CEĞİM	1469	NUNDA	1333	MIŞLA	1232	UĞUMU	1167	ELİND	1118
AKŞAM	1579	HABER	1469	DILAR	1332	ULLAN	1231	KİTAB	1166	LİŞKİ	1117
MLERİ	1572	YORUZ	1468	KULLA	1329	TIYOR	1230	ÇATLI	1165	HATTA	1116
IĞIND	1572	NLARA	1467	GÜNLE	1329	KINDA	1230	MEMİŞ	1164	HEPSİ	1114
KARAN	1567	DINLA	1464	İRİNİ	1327	EĞİLD	1225	İÇERİ	1162	DEMİŞ	1109
KANLI	1565	LARIY	1461	ORMUŞ	1326	ĞİLDİ	1223	YAZAR	1162	ERDİM	1106
MAMIŞ	1564	YAKIN	1455	TIĞIN	1326	İSİNE	1220	ONUND	1160	MERAK	1106
SABAH	1557	İLDİĞ	1453	LATTI	1325	ARISI	1219	KANIN	1160	EVLET	1104
YACAK	1550	NELER	1452	ARDIR	1324	LENME	1217	GİRDİ	1158	BENDE	1102
EYLER	1549	IYORL	1450	EREDE	1324	YAKLA	1215	HAZIR	1158	ACAKL	1102
ABİLİ	1548	ÜNLER	1449	ENDEN	1320	SİZLİ	1214	IZLAR	1157	PARÇA	1102
LADIĞ	1547	ATLAR	1448	AMANL	1319	ANNEM	1214	DAŞLA	1157	GÖREV	1101
RIYLA	1546	ASKER	1443	DİNLE	1316	İRMİŞ	1211	ELERD	1156	ALABA	1098
ARDIM	1545	POLİS	1443	LERİY	1312	HANGİ	1209	BİRLE	1155	BALAR	1097
SANLA	1542	HİKAY	1442	ĞİMİZ	1312	MAYAN	1206	HANIM	1155	APLAR	1096
TILAR	1541	HAYAL	1438	RADAN	1312	CAKTI	1205	SİLAH	1152	İSSET	1092

2.1.3. Affine Cipher

The encryption method aims to perform the encryption process on the linear equation of $y_i = ax_i+b \pmod n$ where x_i refers to plaintext, y_i refers to ciphertext, n refers to size of the alphabet and pair of (a, b) refer to key. In the equation, (a, n) must be coprime. Substitution cipher is a special form of Affine cipher where a is equal to 1. And vice versa, affine cipher is a special form of substitution cipher where composing substitution table is formulised.

Table 2.8 Index Values of the Letter of the Turkish Alphabet

A	B	C	Ç	D	E	F	G	Ğ	H	I	İ	J	K	L	M	N	O	Ö	P	R	S	Ş	T	U	Ü	V	Y	Z
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28

Suppose that the plaintext is *ALANKAY* and the key is $(2, -4)$[2]. The plaintext would have the corresponding values found in Table 2.8. By using the plaintext, it is encrypted as illustrated in Table 2.9 and the ciphertext is decrypted as demonstrated in Table 2.10.

Table 2.9 Encryption Process of Affine Cipher

Plaintext	*A*	*L*	*A*	*N*	*K*	*A*	*Y*
Index values	0	14	0	16	13	0	27
y= 2x - 4	0.2-4	14.2-4	0.2-4	16.2-4	13.2-4	0.2-4	27.2-4
y= 2x - 4	-4	24	-4	28	22	-4	50
Mod 29	25	24	25	28	22	25	21
Ciphertext	*Ü*	*U*	*Ü*	*Z*	*Ş*	*Ü*	*S*

Table 2.10 Decryption Process of Affine Cipher

Ciphertext	*Ü*	*U*	*Ü*	*Z*	*Ş*	*Ü*	*S*
Index values	25	24	25	28	22	25	21
x=(y+4)/2	(25+4)/2	(24+4)/2	(25+4)/2	(28+4)/2	(22+4)/2	(25+4)/2	(21+4)/2
x=(y+4)/2	14,5	14	14,5	16	13	14,5	12,5
Mod 29	$\frac{145+5.29}{10}$	14	$\frac{145+5.29}{10}$	16	13	$\frac{145+5.29}{10}$	$\frac{125+5.29}{10}$
Mod 29	0	14	0	16	13	0	27
Plaintext	*A*	*L*	*A*	*N*	*K*	*A*	*Y*

[2] The key (2, -4) is equal to (2, 25) because of the principle y mod n = ax+b mod n = ax mod n + b mod n

The key space of the affine cipher is 812 for Turkish[3]. Therefore, the method is weak against to brute force attack. Moreover, the method is also insecure against to frequency analysis attack because the method is a special form of substitution cipher. Furthermore, if two ciphertext symbols of the plaintext are detected, the key could be calculated from the equation easily.

2.2. Block Ciphers

Most generally, plaintext is divided into blocks, and each block is encrypted respectively in block ciphers. Block ciphers manipulates the frequency distribution of the ciphertext and makes the frequency analysis attacks difficult.

2.2.1. Permutation Cipher

Permutation cipher depends on the principle that transposing the plaintext letters each other. In order to implement the method, a permutation rule is specified and the plaintext divided into blocks. Finally, each block is permuted within the permutation rule.

Figure 2.3 illustrates an example of permutation cipher.

The method is weak against to frequency analysis attack because letters of the plaintext would not be changed in the ciphertext. Moreover, the key space of the permutation cipher is equal to m! and m could be equal to the length of the plaintext in worst case. Therefore, the method could be revealed by brute force attack easily.

[3] The number of letters of the Turkish alphabet is equal to 29 which is a prime number. In the equation y= ax+b mod n, a has to be coprime with 29 to equation be reversible. That's why, key space of the a is equal to 28 according to restriction. Moreover, key space of the b is always equal to n. Thus, key space of the affine cipher 28x29=812 for Turkish.

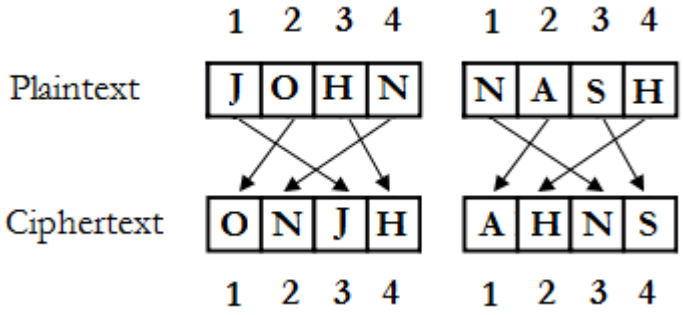

Figure 2.3 Illustration of the Permutation Cipher

2.2.2. Polygraphic Substitution

Polygraphic substitution performs substitutions on blocks instead of substituting a letter. Substituting multiple characters at a time destroys the structure of the plaintext. That makes the frequency analysis attacks impossible.

2.2.2.1. Playfair Cipher

Playfair cipher is one of the best known encryption techniques of multi-character substitution. Firstly, the encryption key is picked up. Secondly, the key would be filled into the 5x5 matrix. However, duplicated letters has to be dropped in the key. Remaining letters in the alphabet would be filled into the rest of the cells of matrix alphabeticlly. The letters of the alphabet has to be filled into 25 cells. Therefore, some letters can be filled into same cells.

Encryption rule is based on a simple technique. The ciphertext is grouped into blocks of length 2. The encryption rule performs on each block. Specific letters should be inserted between repeating letters in the plaintext. If the letters of the current block are located in the same row of the matrix, the current block is replaced by letters to the right of each letter in the row. If the letters of the current block are located in the same column, the current block is replaced by letters below them. Otherwise, letters of the current block are located in neither same row nor same column, a virtual rectangle is composed. The rectangle has to contain the letters of the current block in the corner. The block is replaced by the the other corner elements.

Suppose that the key is *İstanbul* and the plaintext is *Florya*. The key is filled into the matrix as illustrated in Table 2.11. The plaintext is grouped into blocks of length 2.

FL OR YA

Table 2.11 The Encryption Key of Playfair Cipher

İ	S	T	A	N
B	U	L	C/Ç	D
E	F	G/Ğ	H	I/J
K	M/N	O/Ö	P	R
Ş	Ü	V	Y	Z

Finally, the ciphertext is represented as:

GU PK AC

Thus, the plaintext *FLORYA* is encrypted as *GUPKAC* in Playfair Cipher.

Frequency analysis attacks fail against to playfair cipher [4]. It is a fact that the playfair cipher is based on bigram substitution. Therefore, the cipher would not balance out the bigram frequencies of ciphertext. Bigram frequencies of the source language could be used to attack playfair ciphers.

2.2.2.2. Hill Cipher

Hill Cipher is a block cipher model based on matrix manipulations. The plaintext is divided into blocks and each block would be encrypted respectively. Suppose that block size is decided as n. A (nxn) matrix is specified as key and the each block of the plaintext is multiply to the matrix to encrypt the plaintext.

$$\vec{C}=[K]\,\vec{P} \bmod m \qquad (2.3)$$

C and P are column vectors of length n, representing the plaintext and ciphertext respectively, and K is a nxn matrix, which is the encryption key. Decryption requires using the inverse matrix of the matrix K.

$$\vec{P}=[K]^{-1}\vec{C} \bmod m \tag{2.4}$$

The inverse matrix K^{-1} is defined by the equation $K\,K^{-1} = I$, where I is the Identity matrix. It is a fact that the inverse matrix does not always exist.

Suppose that the size of the key matrix is 2x2, and the key matrix is

$$K = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$$

Assume the plaintext would be *alankay*. First, the plaintext is grouped into blocks of length 2. The first letter of the plaintext is appended at the end of the plaintext because the plaintext consists of odd number.

al *an* *ka* *ya*

Secondly, letters of the plaintext are replaced by the corresponding numbers

0 14 0 16 13 0 27 0

Thirdly, each block is multiplied by K

$$\begin{bmatrix} 0 \\ 14 \end{bmatrix}\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} 0.1 + 14.2 \\ 0.3 + 14.4 \end{bmatrix} = \begin{bmatrix} 28 \\ 56 \end{bmatrix} mod\ 29 = \begin{bmatrix} 28 \\ 27 \end{bmatrix}$$

$$\begin{bmatrix} 0 \\ 16 \end{bmatrix}\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} 0.1 + 16.2 \\ 0.3 + 16.4 \end{bmatrix} = \begin{bmatrix} 32 \\ 64 \end{bmatrix} mod\ 29 = \begin{bmatrix} 3 \\ 6 \end{bmatrix}$$

$$\begin{bmatrix} 13 \\ 0 \end{bmatrix}\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} 13.1 + 0.2 \\ 13.3 + 0.4 \end{bmatrix} = \begin{bmatrix} 13 \\ 39 \end{bmatrix} mod\ 29 = \begin{bmatrix} 13 \\ 10 \end{bmatrix}$$

$$\begin{bmatrix}27\\0\end{bmatrix}\begin{bmatrix}1 & 2\\3 & 4\end{bmatrix} = \begin{bmatrix}27.1 + 0.2\\27.3 + 0.4\end{bmatrix} = \begin{bmatrix}27\\81\end{bmatrix} mod\ 29 = \begin{bmatrix}27\\23\end{bmatrix}$$

Fourth, the numbers

28 27 3 6 13 10 27 23

represented as:

ZYÇFKIYT

Thus, the ciphertext *ALANKAYA* would be encrypted as *ZYÇFKIYT* in Hill Cipher.

The inverse matrix is used to implement the decryption process. A matrix has an inverse matrix if and only if its determinant is not equal to 0.

$$K=\begin{bmatrix}a & b\\c & d\end{bmatrix},\ |K|=ad-bc \qquad (2.5)$$

$$K^{-1}=\frac{1}{|K|}\begin{bmatrix}d & -b\\-c & a\end{bmatrix},\ |K|\neq 0 \qquad (2.6)$$

$$KK^{-1} = \begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}$$

Firstly, the inverse matrix is calculated to decrypt the ciphertext.

$$|K| = 1.4 - 2.3 = -2 \neq 0$$

$$K^{-1} = \frac{1}{-2}\begin{bmatrix}4 & -2\\-3 & 1\end{bmatrix} = \begin{bmatrix}-2 & 1\\1{,}5 & -0{,}5\end{bmatrix} mod\ 29$$

$$K^{-1} = \begin{bmatrix}\frac{27}{15 + 5.29} & \frac{1}{5.29 - 5}\\10 & 10\end{bmatrix} mod 29$$

$$K^{-1} = \begin{bmatrix}27 & 1\\16 & 14\end{bmatrix}$$

Secondly, the letters is grouped into blocks

ZY ÇF KI YT

Thirdly, the letters is replaced by the corresponding numbers

28 27 3 6 13 10 27 23

Fourthly, each block is multiplied by K^{-1}

$$\begin{bmatrix}28\\27\end{bmatrix}\begin{bmatrix}27 & 1\\16 & 14\end{bmatrix} = \begin{bmatrix}28.27 + 27.1\\28.16 + 27.14\end{bmatrix} = \begin{bmatrix}783\\826\end{bmatrix} mod\ 29 = \begin{bmatrix}0\\14\end{bmatrix}$$

$$\begin{bmatrix}3\\6\end{bmatrix}\begin{bmatrix}27 & 1\\16 & 14\end{bmatrix} = \begin{bmatrix}3.27 + 1.6\\3.16 + 6.14\end{bmatrix} = \begin{bmatrix}87\\132\end{bmatrix} mod\ 29 = \begin{bmatrix}0\\16\end{bmatrix}$$

$$\begin{bmatrix}13\\10\end{bmatrix}\begin{bmatrix}27 & 1\\16 & 14\end{bmatrix} = \begin{bmatrix}13.27 + 10.1\\13.16 + 10.14\end{bmatrix} = \begin{bmatrix}361\\348\end{bmatrix} mod\ 29 = \begin{bmatrix}13\\0\end{bmatrix}$$

$$\begin{bmatrix}27\\23\end{bmatrix}\begin{bmatrix}27 & 1\\16 & 14\end{bmatrix} = \begin{bmatrix}27.27 + 23.1\\27.16 + 23.14\end{bmatrix} = \begin{bmatrix}752\\754\end{bmatrix} mod\ 29 = \begin{bmatrix}27\\0\end{bmatrix}$$

Finally, the numbers

0 14 0 16 13 0 27 0

represented as:

alankay

Hill cipher is secure against to frequency analysis attack. Moreover, the Hill cipher is strong against to brute force attack. The key space of the hill cipher is equal to 29^{nxn} which is an efficiently large value. However, Hill cipher succumbs to a known plaintext attack [3]. The key matrix can be calculated easily from a set of known $\vec{P}, \vec{C}$.

2.2.3. Polyalphabetic Substitution Ciphers

Remaining the substitution rules same for every substitution makes the frequency analysis attacks applicable on monoalphabetic ciphers. In a polyalphabetic substitution cipher, multiple substitution alphabets are used throughout the encryption process. Therefore, same plaintext character could be encrypted to different ciphertext symbols. That makes the frequency analysis difficult.

2.2.3.1. Vigenere Cipher

Suppose that an n character alphabet and an m character key, $K = (k_1, k_2, \ldots, k_m)$ is given. The Vigenere cipher consists of m shift ciphers and k_i specifies the monoalphabetic substitution [10].

Encryption process could be implemented by adding plaintext and key values. Accordingly, decryption process could be performed by substruction of ciphertext and key values.

$$C_i = (P_i + k_i) \bmod n \quad (2.7)$$

$$P_i = (C_i - k_i) \bmod n \quad (2.8)$$

Suppose that the plaintext is *MUSTAFAKEMALATATÜRK* and the key is *İSTANBUL*. The plaintext is encrypted as illustrated in Table 2.12 and the ciphertext is decrypted as demonstrated in Table 2.13.

Table 2.12 Mathematical Illustration of Encyrption Process of Vigenere Cipher

Plaintext	M	U	S	T	A	F	A	K	E	M	A	L	A	T	A	T	Ü	R	K
Key	İ	S	T	A	N	B	U	L	İ	S	T	A	N	B	U	L	İ	S	T
Plaintext	15	24	21	23	0	6	0	13	5	15	0	14	0	23	0	23	25	20	13
Key	11	21	23	0	16	1	24	14	11	21	23	0	16	1	24	14	11	21	23
Plaintext+Key	26	45	44	23	16	7	24	27	16	36	23	14	16	24	24	37	36	41	36
mod 29	26	16	15	23	16	7	24	27	16	7	23	14	16	24	24	8	7	12	7
Ciphertext	V	N	M	T	N	G	U	Y	N	G	T	L	N	U	U	Ğ	G	J	G

Table 2.13 Mathematical Illustration of Decryption Process of Vigenere Cipher

Ciphertext	V	N	M	T	N	G	U	Y	N	G	T	L	N	U	U	Ğ	G	J	G
Key	İ	S	T	A	N	B	U	L	İ	S	T	A	N	B	U	L	İ	S	T
Ciphertext	26	16	15	23	16	7	24	27	16	7	23	14	16	24	24	8	7	12	7
Key	11	21	23	0	16	1	24	14	11	21	23	0	16	1	24	14	11	21	23
Ciphertext-Key	15	-5	-8	23	0	6	0	13	5	-14	0	14	0	23	0	-6	-4	-9	-16
mod 29	15	24	21	23	0	6	0	13	5	15	0	14	0	23	0	23	25	20	13
Plaintext	M	U	S	T	A	F	A	K	E	M	A	L	A	T	A	T	Ü	R	K

Table 2.14 Encryption Process of Vigenere Cipher by Using of Vigenere Table

Plaintext	M	U	S	T	A	F	A	K	E	M	A	L	A	T	A	T	Ü	R	K
Key	İ	S	T	A	N	B	U	L	İ	S	T	A	N	B	U	L	İ	S	T
Ciphertext	V	N	M	T	N	G	U	Y	N	G	T	L	N	U	U	Ğ	G	J	G

Table 2.15 Vigenere Table for Turkish Alphabet

	A	B	C	Ç	D	E	F	G	Ğ	H	I	İ	J	K	L	M	N	O	Ö	P	R	S	Ş	T	U	Ü	V	Y	Z
A	A	B	C	Ç	D	E	F	G	Ğ	H	I	İ	J	K	L	M	N	O	Ö	P	R	S	Ş	T	U	Ü	V	Y	Z
B	B	C	Ç	D	E	F	G	Ğ	H	I	İ	J	K	L	M	N	O	Ö	P	R	S	Ş	T	U	Ü	V	Y	Z	A
C	C	Ç	D	E	F	G	Ğ	H	I	İ	J	K	L	M	N	O	Ö	P	R	S	Ş	T	U	Ü	V	Y	Z	A	B
Ç	Ç	D	E	F	G	Ğ	H	I	İ	J	K	L	M	N	O	Ö	P	R	S	Ş	T	U	Ü	V	Y	Z	A	B	C
D	D	E	F	G	Ğ	H	I	İ	J	K	L	M	N	O	Ö	P	R	S	Ş	T	U	Ü	V	Y	Z	A	B	C	Ç
E	E	F	G	Ğ	H	I	İ	J	K	L	M	N	O	Ö	P	R	S	Ş	T	U	Ü	V	Y	Z	A	B	C	Ç	D
F	F	G	Ğ	H	I	İ	J	K	L	M	N	O	Ö	P	R	S	Ş	T	U	Ü	V	Y	Z	A	B	C	Ç	D	E
G	G	Ğ	H	I	İ	J	K	L	M	N	O	Ö	P	R	S	Ş	T	U	Ü	V	Y	Z	A	B	C	Ç	D	E	F
Ğ	Ğ	H	I	İ	J	K	L	M	N	O	Ö	P	R	S	Ş	T	U	Ü	V	Y	Z	A	B	C	Ç	D	E	F	G
H	H	I	İ	J	K	L	M	N	O	Ö	P	R	S	Ş	T	U	Ü	V	Y	Z	A	B	C	Ç	D	E	F	G	Ğ
I	I	İ	J	K	L	M	N	O	Ö	P	R	S	Ş	T	U	Ü	V	Y	Z	A	B	C	Ç	D	E	F	G	Ğ	H
İ	İ	J	K	L	M	N	O	Ö	P	R	S	Ş	T	U	Ü	V	Y	Z	A	B	C	Ç	D	E	F	G	Ğ	H	I
J	J	K	L	M	N	O	Ö	P	R	S	Ş	T	U	Ü	V	Y	Z	A	B	C	Ç	D	E	F	G	Ğ	H	I	İ
K	K	L	M	N	O	Ö	P	R	S	Ş	T	U	Ü	V	Y	Z	A	B	C	Ç	D	E	F	G	Ğ	H	I	İ	J
L	L	M	N	O	Ö	P	R	S	Ş	T	U	Ü	V	Y	Z	A	B	C	Ç	D	E	F	G	Ğ	H	I	İ	J	K
M	M	N	O	Ö	P	R	S	Ş	T	U	Ü	V	Y	Z	A	B	C	Ç	D	E	F	G	Ğ	H	I	İ	J	K	L
N	N	O	Ö	P	R	S	Ş	T	U	Ü	V	Y	Z	A	B	C	Ç	D	E	F	G	Ğ	H	I	İ	J	K	L	M
O	O	Ö	P	R	S	Ş	T	U	Ü	V	Y	Z	A	B	C	Ç	D	E	F	G	Ğ	H	I	İ	J	K	L	M	N
Ö	Ö	P	R	S	Ş	T	U	Ü	V	Y	Z	A	B	C	Ç	D	E	F	G	Ğ	H	I	İ	J	K	L	M	N	O
P	P	R	S	Ş	T	U	Ü	V	Y	Z	A	B	C	Ç	D	E	F	G	Ğ	H	I	İ	J	K	L	M	N	O	Ö
R	R	S	Ş	T	U	Ü	V	Y	Z	A	B	C	Ç	D	E	F	G	Ğ	H	I	İ	J	K	L	M	N	O	Ö	P
S	S	Ş	T	U	Ü	V	Y	Z	A	B	C	Ç	D	E	F	G	Ğ	H	I	İ	J	K	L	M	N	O	Ö	P	R
Ş	Ş	T	U	Ü	V	Y	Z	A	B	C	Ç	D	E	F	G	Ğ	H	I	İ	J	K	L	M	N	O	Ö	P	R	S
T	T	U	Ü	V	Y	Z	A	B	C	Ç	D	E	F	G	Ğ	H	I	İ	J	K	L	M	N	O	Ö	P	R	S	Ş
U	U	Ü	V	Y	Z	A	B	C	Ç	D	E	F	G	Ğ	H	I	İ	J	K	L	M	N	O	Ö	P	R	S	Ş	T
Ü	Ü	V	Y	Z	A	B	C	Ç	D	E	F	G	Ğ	H	I	İ	J	K	L	M	N	O	Ö	P	R	S	Ş	T	U
V	V	Y	Z	A	B	C	Ç	D	E	F	G	Ğ	H	I	İ	J	K	L	M	N	O	Ö	P	R	S	Ş	T	U	Ü
Y	Y	Z	A	B	C	Ç	D	E	F	G	Ğ	H	I	İ	J	K	L	M	N	O	Ö	P	R	S	Ş	T	U	Ü	V
Z	Z	A	B	C	Ç	D	E	F	G	Ğ	H	I	İ	J	K	L	M	N	O	Ö	P	R	S	Ş	T	U	Ü	V	Y

Intersection of the plaintext and the key on Vigenere table states the ciphertext. Similarly, the plaintext could be obtained by performing inverse process. A sample scenario is illustrated in Table 2.14 while Table 2.15 is used as Vigenere table.

Note that *A* would be replaced by *N, U, T, N,* and *U* respectively. The action manipulates the frequency distribution of the ciphertext. Therefore, frequency analysis attacks cannot be performed as easily on Vigenere cipher.

```
// C is an array consisting of elements of ciphertext respectively
// K is an array consisting of elements of key respectively

for from i=0 to K.length by 1
        print 'Block '+i+':'
        for from j=i to C.length by K.size
                print C[j]
        end for
end for
```

Figure 2.4 Pseudocode of Dividing Ciphertext Into Blocks to Attack

Suppose that the length of the key is n. In other words, the ciphertext consists of n monoalphabetic substitution ciphers. Note that, the letters of the plaintext at positions 1, n+1, 2n+1, 3n+1 would be encrypted by the same monoalphabetic substitution cipher. Thereby, the ciphertext could be divided into blocks of key length. Implementing frequency analysis to each block respectively makes frequency analysis attacks possible. Figure 2.4 illustrates the algorithm of dividing ciphertext into blocks.

An attacker needs to check all possibilities of key between length of 2 and 28 if the key size is not known. Thus, the key space of the Vigenere cipher would be larger than affine cipher and substitution cipher.

2.2.3.2. Kasiski Attack

Kasiski attack is a method that tries to determine the key size of the polyalpahabatic substitution cipher. The ciphertext could contain a recurrent pattern if the letters of the plaintext are encrypted by the same sequence of key. It is assumed that the distance between the repeating groups of characters could be related with the key length. Greatest common divisor of the distances of the repetations could give a clue about the key size [11].

Suppose that the following ciphertext is given:

AUYAJAMĞÖSCMİRĞYRRMVEIGIZCTSİALNDIFDARBKMNAOULCRMYEİGİ
ZTCZÖAŞBKRBMASYLÜRİÇUZEEBYYMÜZDYZŞRRKVYAİINOĞKRDAGRÜP
ELAÖLCMEBÖRCNSSVİVAYSGSCZOBOUZUNİLUUERRÖDNŞMSOEGÖNÖÖB
CRYSDVRRBYYİÖORİZHÜRŞSİİJÖYBÖMÜKMJZKNNEFÖYŞEYNVLRŞMVFIF
DULĞYĞRUCKNEATNZIÖORİZ

Table 2.16 Distances Between Repating Characters

Character	Times	Distance
CM	2	102
IFD	2	180
BYY	2	90
ÖORİZ	2	60

In the example, repeating characters *CM*, *IFD*, *BYY* and *ÖORİZ* appear two times in the ciphertext. Distances between the characters are illustrated in the Table 2.16.

The greatest common divisor of the 60, 90, 102 and 180 is 6. Therefore, 6 and its dividers (2, 3) are candidates of the key length. Indeed, the key is specified as *ANKARA* and the plaintext is selected from the following text.

AĞLASAMSESİMİDUYARMISINIZMISRALARIMDADOKUNABİLİRMİSİNİZG
ÖZYAŞLARIMAELLERİNİZLEBİLMEZDİMŞARKILARINBUKADARGÜZELKE
LİMELERİNSEKİFAYETSİZOLDUĞUNUBUDERDEDÜŞMEDENÖNCEBİRYERV
ARBİLİYORUMHERŞEYİSÖYLEMEKMÜMKÜNEPEYCEYAKLAŞMIŞIMDUYU
YORUMANLATAMIYORUM

3. Homophonic Cipher

Homophonic cipher is developed as an alternative to substitution cipher to compose more resistant ciphertexts against to the frequency analysis attacks. Homophonic cipher could be thought as the extended version of substitution cipher [2].

In the classical substitution, each plaintext character is replaced with a corresponding symbol by means of one to one mapping. Homophonic substitution is similar to the classical substitution, except that the mapping is one to many. Each source character is mapped into a set of symbols referred to as homophones [12]. The number of substitutes is proportional to the frequency of the letter in the source language. It can be used to provide randomization. That makes the frequency occurrence of the ciphertext symbols more uniform [3]. The term of homophonic means to sound the same that is an indication of the fact that different symbols refer to same source character. An important attribute of the homophonic coding is that a symbol is picked at random from the set of homophones to represent the given source character in the one to many mapping. Thus, homophonic cipher is transforming a given non-uniformly distributed plaintext into a random uniformly distributed ciphertext [13].

Homophonic substitution is also a cryptographic technique that reduces the redundancy of a message [14]. Homophonic cipher replaces each plaintext letter with different symbols proportional to its frequency rate. Its main goal is to convert the plaintext into a sequence of completely random code symbols. The idea behind homophonic cipher is to balance out the symbol frequencies. The frequency distribution of the ciphertext is manipulated and smoothed. Symbols located in the ciphertext have relatively equal frequencies. Each symbol takes space of about one percent of ciphertext.

Suppose that the source language consists of the alphabet A = {a, b} with probabilities p_a=3/4, p_b=1/4 and the plaintext letters would be substituted according to substitution rule a→{00, 01, 10} and b→{11} as illustrated in Figure 3.1. The message is encrypted at random into one of its homophones with equal probabilities [15]. So, the homophonic cipher provides the encrypted text appearing random.

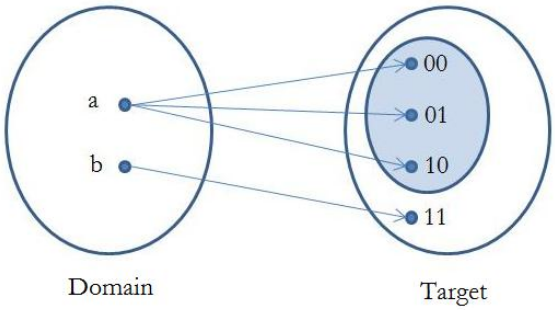

Figure 3.1 Domain Set and Target Set of Homophonic Cipher

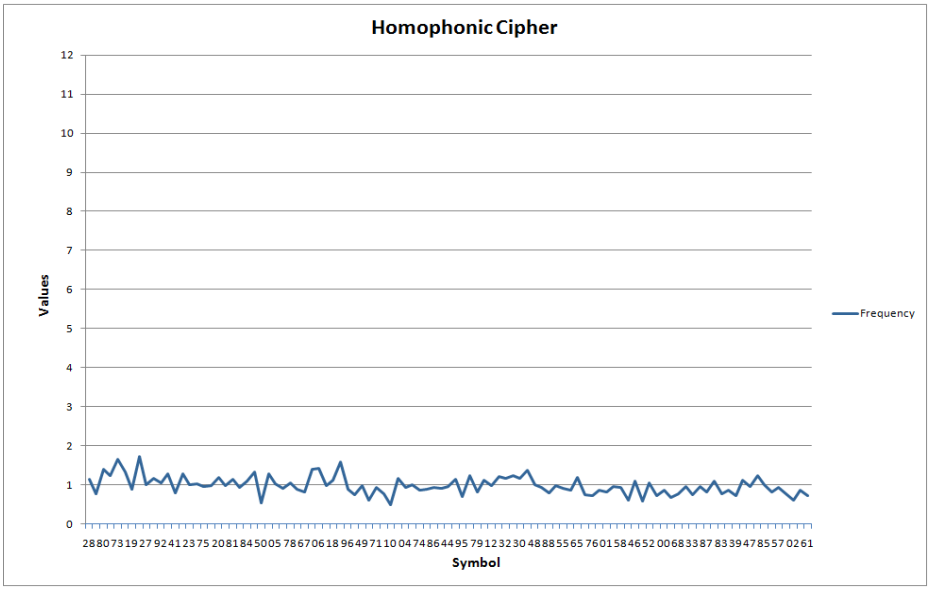

Figure 3.2 Illustration of Frequency Distribution of a Homophonic Encrypted Text

The article of *Bir Tılsımı Olmalıdır Hayatın* written by *Çetin Altan* is encrypted by homophonic cipher corresponding to *Figure 3.3*. The text consists of *3856* letters. Finally, the *Figure 3.2* is retrieved.

Table 3.1 Percentage of Turkish Letter Frequencies and Expression Count in Homophonic Cipher [2]

Letter	Freq.	Exp. Count	Letter	Freq.	Exp. Count	Letter	Freq.	Exp. Count
A	11,92	12	I	5,114	5	R	6,722	7
B	2,844	3	İ	8,6	9	S	3,014	3
C	0,963	1	J	0,034	1	Ş	1,78	2
Ç	1,156	1	K	4,683	5	T	3,314	3
D	4,706	5	L	5,922	6	U	3,235	3
E	8,912	9	M	3,752	4	Ü	1,854	2
F	0,461	1	N	7,484	7	V	0,959	1
G	1,253	1	O	2,476	2	Y	3,336	3
Ğ	1,125	1	Ö	0,777	1	Z	1,5	2
H	1,212	1	P	0,886	1			

The unigram frequencies of the source language assess how many symbols the letter would be expressed within homophonic cipher. Each letter would be replaced by different symbols proportional to its frequency rate. Table 3.1 illustrates the expression count of Turkish unigrams.

A	B	C	Ç	D	E	F	G	Ğ	H	I	İ	J	K	L	M	N	O	Ö	P	R	S	Ş	T	U	Ü	V	Y	Z
009	048	013	062	001	014	010	006	025	023	032	083	015	004	026	022	018	000	072	038	029	011	076	017	008	063	034	021	002
012	081			003	016					070	088		096	037	027	058	005			035	019	086	020	061	085		052	069
033	028			045	024					073	093		056	051	039	059				040	036		030	097			075	
047				079	044					031	060		065	084	050	066				042								
053				041	046					089	007		068	043		071				077								
067					055						054			049		091				080								
078					057						090					101				102								
092					064						099																	
082					074						095																	
087																												
098																												
094																												

Figure 3.3 Substitution Table of Homophonic Cipher

Figure 3.3 illustrates an instance of the substitution table of homophonic cipher. The top row represents unigrams of the Turkish, while the below row represents homophones of each unigram.

Suppose that the plaintext is *FLORYA*. The plaintext could be encrypted as one of the following ciphertexts:

010	*084*	*000*	*035*	*075*	*067*
010	*026*	*005*	*042*	*052*	*098*
010	*043*	*000*	*029*	*021*	*053*

Homophonic cipher is a type of monoalphabetic substitution cipher. A letter could be shown as several characters. However, a character could symbolize only one plaintext character. In polyalphabetic cipher, a letter could be represented by several characters and a character would symbolize several letters throughout encryption. In other perspective, the encryption alphabet remains constant throughout the encryption process [16].

```
exp sym[29]=12, 3, 1, 1, 5, 9, 1, 1, 1, 1, 5, 9, 1, 5, 6, 4, 7, 2, 1, 1, 7, 3, 2, 3, 3, 2, 1, 3, 2
// Expression count of each letter in Turkish Alphabet
num_of_sym=102 // Sum of the expression count of Turkish letters
keyspace=1
b=sym[0]
for from i=0 to 29
        keyspace=keyspace*Combination(num_of_sym, b)
        num_of_sym=num_of_sym-b
        b=sym[i+1]
end for
return keyspace

function Combination (a, b)
dividend=1,  denominator=1
for from i=a downto b by -1
        dividend = dividend * i
end for
for from i=b downto 1 by -1
        denominator = denominator*b
end for
resp = dividend / denominator
return resp
```

Figure 3.4 Algorithm for Computing the Key Space of Homophonic Cipher for Turkish

The key space of homophonic cipher for Turkish is calculated in Figure 3.4. It is a number larger than 10^{119}. Suppose that the attacker is able to check a possibility per microsecond, it would take time more than 10^{105} years to solve ciphertext in worst case[4]. Obviously, the encipherment rules out a brute force attack.

It is understood that brute force attacks would not be a solution to break homophonic cipher if the key space values of well known algorithms are compared. The key space of the method is overwhelmingly larger than other common algorithms, even most of modern algorithms, except Blowfish.

Table 3.2 Comparison of Key Space Values of Common Algorithms

Algorithm	**Category**	**Key Space**	**Approximate Value**
Substitution Cipher	Classical	29!	10^{30}
Homophonic Cipher	Classical	10^{119}	10^{119}
DES	Modern	2^{56}	10^{16}
3DES	Modern	2^{112}	10^{33}
IDEA	Modern	2^{128}	10^{38}
AES	Modern	2^{256}	10^{77}
Camellia	Modern	2^{256}	10^{77}
Twofish	Modern	2^{256}	10^{77}
Serpent	Modern	2^{256}	10^{77}
Blowfish	Modern	2^{448}	10^{134}

In order to compare the key space of common encryption algorithms, *BigDecimal* class of Java programming language is used. Thus, approximate key space values are retrieved illustrated in *Table 3.2*.

[4] $\frac{10^{119}\,10^{-6}}{60.60.24.365} > 10^{105}$

4. A New Approach On Attacking Homophonic Cipher

Most frequent n-grams would not help to solve homophonic ciphers. Even if these n-grams are assumed to appear in ciphertext, they would almost be impossible to solve because of the high expression count.

Homophonic cipher extends the block size of the ciphertext in patches. Moreover, the block size is variable. That makes the identifying the plaintext difficult. Well known statistical features of the source language become invalid because of the manipulation of the ciphetext.

The 100 most common words of Turkish are already illustrated in the section 2. Expression count of the most common words could contribute to solve homophonic cipher. Table 4.1 illustrates the expression count of most common 100 words of Turkish. It was sorted with respect to the expression count from smallest to greatest.

Nevertheless, making a decision of useful n-grams belonging to the source language plays pivotal role to solve homophonic ciphers. The unigrams of n-grams should have low frequencies to be detected easily in homophonic encrypted texts, whereas the n-gram itself should have high frequency to be assumed to appear in the plaintext. In other words, high frequent n-grams should consist of low frequent unigrams [2].

Table 4.1 Expression Count of Most Common 100 Words of Turkish within Homophonic Cipher

Word	**Exp**	**Word**	**Exp**	**Word**	**Exp**	**Word**	**Exp**
O	2	İn	63	İse	243	Dedi	2025
Şu	6	Önce	63	Nın	245	Vardı	2100
Ve	9	Göre	63	Bütün	252	Fakat	2160
Bu	9	Bey	81	Olur	252	Hemen	2268
Hiç	9	Gece	81	İlk	270	Değil	2430
Çok	10	Var	84	Onun	294	Adam	2880
Gün	14	Yıl	90	İki	405	Bana	3024
Mı	20	Küçük	100	Nin	441	Şimdi	3240
Yok	30	Uzun	126	İle	486	Sonra	3528
Çocuk	30	Tek	135	Böyle	486	Karşı	4200
In	35	Çünkü	140	İşte	486	Başka	4320
Ya	36	Tam	144	Olduğu	540	Biraz	4536
Mi	36	Öyle	162	İçin	567	Ancak	5040
Hem	36	Ona	168	Ama	576	Artık	6300
Onu	42	Büyük	180	Daha	720	Benim	6804
Son	42	Oldu	180	Olan	1008	Nasıl	7560
De	45	Bir	189	Diye	1215	Zaman	8064
Ki	45	Ben	189	Eski	1215	Kadın	10500
Kız	50	Sen	189	Aynı	1260	Olduğunu	11340
Şey	54	Bunu	189	Bile	1458	Kendi	14175
Biz	54	Doğru	210	Beni	1701	İnsan	15876
Da	60	Türk	210	Yeni	1701	Kadar	25200
Ne	63	Güzel	216	Yine	1701	İçinde	25515
Her	63	Gibi	243	Biri	1701	Türkiye	51030
En	63	İyi	243	Bizim	1944	Olarak	60480

4.1. High Frequent N-grams Consisting of Low Frequent Unigrams

Turkish n-gram frequencies are initially explored in a large corpus of size 13.4 MB to obtain high frequent n-grams consisting of low frequent unigrams. Secondly, expression count of each line within homophonic cipher is computed. Thirdly, sorting was done with respect to the frequeny values by taking into account the first 300 records for bigrams, 1500 results for trigrams, 2500 results for tetragrams and pentagrams from the greatest to smallest and the rest of data was discarded. Finally, it was sorted with respect to the expression count from the smallest to greatest. N-gram frequencies indicate frequencies in *11.371.564.*

Table 4.2 Frequencies and Expression Count of High Frequent Bigrams Consisting of Low Frequent Unigrams in 11M

Ngr	Fre	Exp	Ngr	Fre	Exp	Ngr	Fre	Exp	Ngr	Fre	Exp	Ngr	Fre	Exp
GÖ	25203	1	ÜT	7852	6	PE	10171	9	LO	4242	12	NT	8977	21
GÜ	20124	2	SÜ	7729	6	US	10074	9	SM	4030	12	YR	5622	21
ÇO	14880	2	OT	7353	6	EÇ	8691	9	OR	70300	14	YN	5457	21
ÖZ	12477	2	PL	7068	6	FE	8636	9	ON	45874	14	RB	5141	21
OĞ	10648	2	ĞL	7029	6	EF	7925	9	ÜN	43059	14	NB	4482	21
OC	7324	2	ÖL	6801	6	EH	7137	9	ÜR	27057	14	TR	4288	21
ÜÇ	5469	2	LG	6372	6	İH	6744	9	RÜ	13654	14	AŞ	46294	24
ÇÜ	5186	2	ŞU	6102	6	Fİ	6520	9	NÜ	10971	14	AZ	32755	24
OP	5075	2	ÇL	5262	6	EP	6092	9	RŞ	5193	14	LM	32197	24
PO	4765	2	ZU	5205	6	İF	5528	9	RO	4899	14	ZA	23774	24
ĞÜ	4516	2	OS	4774	6	SS	4284	9	DU	52283	15	ŞA	22239	24
ÜĞ	4463	2	VL	4745	6	İC	3936	9	TI	51386	15	ML	12350	24
ĞU	18753	3	NC	21614	7	Pİ	3837	9	SI	50362	15	DI	62359	25
UĞ	16907	3	ÖR	17778	7	IŞ	32622	10	KT	28964	15	IK	40865	25
ÖY	14744	3	ÖN	15278	7	OK	27389	10	IY	27516	15	KI	31841	25
SÖ	10196	3	ĞR	7347	7	KO	25764	10	YI	24913	15	KK	5957	25
CU	9701	3	RG	6210	7	DÜ	25363	10	KU	23617	15	ID	4104	25
PT	5061	3	NG	5692	7	IZ	23091	10	UK	15139	15	Bİ	125824	27
UP	4701	3	RÇ	4136	7	ŞI	22692	10	YD	13222	15	İY	57222	27
BÖ	4683	3	VR	3925	7	ZI	14561	10	IS	10610	15	Sİ	55239	27
GU	4374	3	NÇ	3916	7	ÜK	13284	10	KS	7683	15	YE	55184	27
VU	3907	3	MÜ	11630	8	DO	12350	10	SK	6723	15	TE	53881	27
ÜZ	17636	4	ÜM	11563	8	ŞK	11863	10	BI	5760	15	Tİ	50968	27
ÜŞ	13030	4	ŞM	9568	8	KÜ	11279	10	UD	3968	15	ET	49927	27
ZÜ	7025	4	OM	5460	8	ZD	8012	10	İŞ	38227	18	BE	45640	27
ŞÜ	6549	4	VE	49863	9	OD	4565	10	İZ	35238	18	SE	44712	27
OŞ	5093	4	BU	44624	9	HA	56075	12	YL	32373	18	EY	44573	27
ĞI	37718	5	GE	40841	9	OL	54389	12	UL	29371	18	ES	39387	27
IĞ	24106	5	CE	37156	9	CA	33993	12	LU	27338	18	İS	37263	27
ÇI	18112	5	Ğİ	36283	9	AH	33171	12	ZE	22563	18	İT	20510	27
CI	10287	5	Gİ	35787	9	AP	27085	12	ŞE	19742	18	Yİ	19854	27
IP	10041	5	ST	31918	9	AĞ	26617	12	TL	19492	18	İB	15851	27
PI	8685	5	İÇ	31429	9	UM	26139	12	Şİ	17282	18	EB	5667	27
DÖ	6918	5	Çİ	27377	9	VA	23616	12	Zİ	13862	18	RM	24077	28
KÖ	5975	5	EV	25568	9	MU	23551	12	EŞ	13465	18	NM	13100	28
FI	4036	5	TU	22533	9	PA	21525	12	EZ	9758	18	LI	53814	30
HI	3786	5	SU	22168	9	AC	18985	12	SL	7774	18	KL	43444	30
YO	61044	6	İĞ	21975	9	ŞL	18154	12	LT	7738	18	IL	39898	30
SO	27989	6	HE	21182	9	ÇA	17193	12	LS	4266	18	LD	37375	30
ŞT	26972	6	UY	20367	9	ZL	16237	12	IM	37245	20	LK	9755	30
TÜ	23620	6	EC	19480	9	TM	13595	12	MI	36784	20	IN	132807	35
UZ	15172	6	EĞ	19472	9	AV	13357	12	MD	11337	20	ND	98136	35
UŞ	14641	6	TT	19323	9	LÜ	13272	12	KM	7097	20	RD	71256	35
YÜ	14156	6	Hİ	18821	9	ÜL	13181	12	UN	59182	21	NI	69437	35
BÜ	11348	6	ÇE	17736	9	AÇ	12668	12	UR	38171	21	RI	66743	35
ÜY	11253	6	UT	13198	9	FA	11877	12	NU	36391	21	IR	43789	35
ÜS	10146	6	Cİ	11387	9	AF	11150	12	RU	28453	21	RK	29528	35
TO	9312	6	YU	11219	9	GA	10575	12	RT	20288	21	NK	7820	35
BO	9184	6	İP	10803	9	ĞA	6727	12	RS	15816	21	KR	3777	35
OY	8640	6	Vİ	10524	9	MS	4790	12	NS	9885	21	YA	107833	36

Table 4.3 Frequencies and Expression Count of High Frequent Trigrams Consisting of Low Frequent Unigrams in 11M

Ngr	Freq	Exp	Ngr	Freq	Exp	Ngr	Freq	Exp	Ngr	Freq	Exp	Ngr	Freq	Exp
GÖZ	5755	2	ÇOK	9685	10	CUK	3185	15	ÜKÜ	1016	20	ÇTİ	1610	27
ÇOC	3833	2	DOĞ	4452	10	YIP	1715	15	ĞUN	7798	21	PSİ	1575	27
GÜV	1092	2	DÜĞ	2915	10	PIY	1561	15	ĞRU	2722	21	HTİ	1488	27
HOC	1017	2	KÜÇ	2321	10	TIP	1414	15	UNC	1857	21	GİY	1407	27
GÖS	2247	3	ÇÜK	1907	10	KÖY	1288	15	RUP	1662	21	EPS	1338	27
GÖT	1143	3	KOC	1906	10	KÖT	1010	15	VUR	1636	21	HEY	1311	27
ÜĞÜ	4160	4	KÖŞ	1290	10	MÜŞ	2801	16	ÖRT	1431	21	TİH	1305	27
ÖZÜ	2872	4	IZC	1094	10	ÜMÜ	2380	16	RGU	1275	21	EÇT	1248	27
ÜÇÜ	2810	4	HIZ	1063	10	ÜŞM	1709	16	YÖN	1235	21	CES	1232	27
GÜZ	2803	4	ÜTÜ	5629	12	MÜZ	1447	16	GUN	1102	21	SUS	1188	27
ÜCÜ	1607	4	YÜZ	5561	12	ZÜM	959	16	VRU	975	21	ÇET	1150	27
HOŞ	1540	4	CAĞ	4672	12	ÖYL	13810	18	MUŞ	6665	24	ÖTE	1114	27
KÖP	1250	5	ÜYO	4299	12	UYO	5892	18	OCA	3344	24	FET	1101	27
OCU	3910	6	ÜYÜ	4106	12	ÜST	5059	18	GAZ	2805	24	SEF	1070	27
OĞU	3155	6	ÖZL	3266	12	BÜT	4849	18	ÜLÜ	2684	24	TTU	1069	27
TOP	2979	6	GÜL	2851	12	ŞTU	3803	18	HAZ	2055	24	UTT	1065	27
SÖZ	2686	6	ĞUM	2715	12	UZU	3780	18	FAZ	1790	24	GİS	1015	27
ÖTÜ	2203	6	HAF	2525	12	UŞT	3615	18	VAŞ	1767	24	TEC	1008	27
FÜS	1824	6	ÖLÜ	2404	12	BÜY	3577	18	ŞAH	1684	24	ÜRÜ	6600	28
ÖLG	1038	6	OĞL	2295	12	OYU	3519	18	UŞM	1674	24	ÜNÜ	6358	28
BOĞ	997	6	POL	2079	12	OTU	3383	18	MUZ	1672	24	ŞÜN	4498	28
ŞÖY	982	6	OTO	2061	12	ÖZE	2459	18	OĞA	1475	24	ZÜN	3283	28
GÖR	12199	7	HAV	2005	12	STÜ	2345	18	OMU	1287	24	ÖRM	2011	28
ÖNC	4016	7	BOŞ	1984	12	ŞEH	2272	18	ZCA	1195	24	ZOR	1800	28
ÖĞR	2452	7	ÜŞT	1937	12	LUĞ	2258	18	ŞAĞ	1139	24	RÜŞ	1040	28
GÖN	1184	7	OPL	1739	12	BOY	2202	18	ÜMS	1077	24	IYO	15455	30
ÜŞÜ	5781	8	ŞTÜ	1615	12	UŞU	1990	18	APO	1062	24	ŞTI	10166	30
ÜZÜ	3851	8	ÇAĞ	1244	12	SUZ	1927	18	MÜS	1026	24	IŞT	8101	30
ĞÜM	964	8	MUH	1233	12	ĞLU	1873	18	TÜM	955	24	OKU	6541	30
UĞU	15363	9	AHV	1213	12	ÜSU	1827	18	IĞI	22842	25	LIĞ	6354	30
GEÇ	7184	9	ÜSÜ	1051	12	ĞİŞ	1520	18	ÇIK	11312	25	YOK	5827	30
HİÇ	7076	9	AHÇ	1029	12	GİZ	1499	18	DIĞ	11092	25	YÜK	4374	30
SÖY	6645	9	CUM	1026	12	EFO	1372	18	ICI	1863	25	SIZ	3789	30
CEĞ	4410	9	VAP	978	12	ÜVE	1366	18	KIP	1833	25	OKT	2768	30
HEP	3254	9	LÜĞ	968	12	OST	1336	18	GİB	11642	27	SOK	2715	30
GEC	3087	9	GÜN	8955	14	BÖL	1255	18	SEV	5934	27	PIL	2599	30
BÖY	3016	9	ÖRÜ	4600	14	SYO	1195	18	USU	5550	27	ÇIL	2183	30
UCU	2433	9	ÖNÜ	3537	14	GEZ	1141	18	GİT	4392	27	ŞIY	2095	30
ÖST	2287	9	OĞR	3298	14	ÖŞE	1115	18	UYU	4351	27	ÖLD	1930	30
UYG	1996	9	ÇÜN	2195	14	GUL	1084	18	TİĞ	4325	27	KOY	1928	30
YGU	1940	9	ÜNC	2063	14	UÇL	1080	18	YEC	3166	27	CIL	1705	30
ÇEV	1737	9	ĞÜN	1837	14	VİZ	1025	18	TUT	3059	27	LIP	1554	30
HÇE	1719	9	ORG	1560	14	LUP	1023	18	UTU	2586	27	OKS	1389	30
CEV	1278	9	FON	1444	14	CUL	987	18	VET	2387	27	KUŞ	1326	30
SUÇ	1217	9	POR	1429	14	DÜŞ	6913	20	ÇBİ	2379	27	ÜKS	1317	30
TUĞ	1215	9	GÜR	1253	14	ĞIM	6633	20	İÇB	2369	27	DOS	1195	30
HVE	1090	9	RGÜ	1218	14	ÜDÜ	2013	20	HİS	2271	27	ĞLI	1169	30
ÖPE	1043	9	DUĞ	10666	15	DÜZ	1357	20	GET	2144	27	TIŞ	935	30
EVG	962	9	TIĞ	4856	15	KOŞ	1151	20	UST	1880	27	ĞIN	12555	35
VGİ	954	9	PTI	3326	15	ÜZD	1104	20	İHT	1641	27	DÖN	4583	35

Table 4.4 Frequencies and Expression Count of High Frequent Tetragrams Consisting of Low Frequent Unigrams in 11M

Ngr	Fre	Exp	Ngr	Fre	Exp	Ngr	Fre	Exp	Ngr	Fre	Exp
GÖZÜ	1760	4	GÜNÜ	1046	28	ÖBÜR	554	42	ÇEVR	804	63
GÜCÜ	628	4	ĞÜNÜ	915	28	DUĞU	10666	45	URUP	767	63
ÇOCU	3833	6	ÖRÜŞ	893	28	DUYG	1360	45	VURU	665	63
GÖTÜ	1132	6	ÖZÜN	831	28	TTIĞ	1069	45	UCUN	587	63
OCUĞ	743	6	ŞÜNC	679	28	KÖPE	899	45	HERH	570	63
ÇOĞU	692	6	ÖNÜŞ	611	28	ÖFKE	843	45	RUCU	563	63
GÖST	2246	9	OCUK	3087	30	ÖPEK	737	45	ÖRDÜ	2573	70
CUĞU	699	9	KÖTÜ	1007	30	KTUP	580	45	DOĞR	2562	70
GÖZL	2267	12	DOĞU	1002	30	VDİĞ	569	45	ÇÜNK	1786	70
FOTO	732	12	PIYO	989	30	YDUĞ	527	45	DÖNÜ	1394	70
OTOĞ	683	12	ŞTIĞ	915	30	ÖLÜM	1533	48	RDÜĞ	1261	70
SÖZÜ	633	12	OĞUK	595	30	MÜŞT	1046	48	ÖNDÜ	1085	70
CUMH	613	12	ÜMÜZ	602	32	OĞAZ	635	48	GÜND	902	70
GÜÇL	583	12	GÖRD	2411	35	ŞIĞI	533	50	NIZC	814	70
GÖRÜ	4073	14	GÖND	814	35	SÖYL	6638	54	MUŞT	1731	72
ÖRGÜ	909	14	BÜTÜ	4768	36	UŞTU	3451	54	LÜYO	1131	72
PTIĞ	1419	15	BÜYÜ	3575	36	BÖYL	3012	54	ÇMİŞ	966	72
KUVV	574	15	GÜZE	2774	36	LUĞU	2046	54	UMUZ	942	72
ÜĞÜM	962	16	UĞUM	2499	36	BOYU	1050	54	OMUT	780	72
ÖZÜM	608	16	ÜSTÜ	2298	36	ÖLGE	1034	54	OCAS	746	72
FÜSU	1797	18	OĞLU	1864	36	ĞİŞT	814	54	BOĞA	700	72
GÜVE	1092	18	TOPL	1673	36	SUÇL	767	54	YÜZL	697	72
BUGÜ	803	18	GEÇM	1436	36	VİZY	750	54	MÜST	684	72
HUZU	764	18	ĞUMU	1222	36	ULUĞ	634	54	ÜLTÜ	649	72
SOĞU	678	18	ŞÖYL	982	36	VGİL	628	54	MUYO	584	72
BÖLG	600	18	ÖYLÜ	953	36	YGUL	575	54	OLUŞ	560	72
DÜĞÜ	2719	20	BAHÇ	943	36	ÜŞÜN	4495	56	UMUŞ	524	72
KÜÇÜ	2320	20	UHAF	935	36	ÜZÜN	2446	56	TIĞI	4806	75
ÜÇÜK	1876	20	OLCU	794	36	ŞÜNÜ	1264	56	ÇIKT	1764	75
ÖZÜK	595	20	HAFT	720	36	ZÜNÜ	829	56	KTIĞ	999	75
VRUP	609	21	SÖZL	677	36	CAĞI	4623	60	PISI	903	75
YÜZÜ	2595	24	UMHU	613	36	ÜYÜK	2902	60	TIPK	727	75
ÜŞTÜ	1607	24	HAYV	608	36	ÖLDÜ	1851	60	CISI	668	75
GÜLÜ	1314	24	BÖLÜ	593	36	ŞIYO	1508	60	DÜŞM	1061	80
HOCA	1017	24	YOLC	574	36	KAHV	1141	60	MÜDÜ	970	80
LÜĞÜ	887	24	OLUP	568	36	HIZL	953	60	ŞMIŞ	879	80
HİÇB	2367	27	OTOB	559	36	ÇAĞI	827	60	CEĞİ	4351	81
UYGU	1915	27	TOBÜ	548	36	YÜZD	732	60	ECEĞ	4324	81
GEÇT	1131	27	OBÜS	548	36	ZLIĞ	544	60	GECE	3002	81
HEPS	1119	27	DÜŞÜ	4740	40	ÇAKÇ	535	60	TTİĞ	2865	81
TUĞU	1036	27	PMIŞ	567	40	UĞUN	7229	63	GEÇİ	2546	81
SEVG	955	27	OĞRU	2540	42	ĞUNU	5384	63	GİTT	2324	81
HÇET	807	27	ÖTÜR	1182	42	ÖNCE	3964	63	ÖSTE	2286	81
BEHÇ	807	27	RGÜT	802	42	GÖRE	3399	63	HİSS	2062	81
YGUS	634	27	UGÜN	801	42	ÖĞRE	2436	63	GEÇE	1655	81
GUSU	600	27	ORGU	784	42	GERÇ	1926	63	TUTU	1144	81
UVVE	549	27	TOĞR	680	42	GENÇ	1646	63	UTTU	963	81
VVET	546	27	OĞUN	562	42	ÖREV	1101	63	HTİY	956	81
GÖRM	1985	28	ÖRÜY	560	42	PENC	1072	63	EVGİ	954	81
ÜĞÜN	1798	28	PORT	560	42	YUNC	1005	63	EHÇE	813	81

Table 4.5 Frequencies and Expression Count of High Frequent Pentagrams Consisting of Low Frequent Unigrams in 11M

Ngr	Fre	Exp	Ngr	Fre	Exp	Ngr	Fre	Exp	Ngr	Fre	Exp
ÇOCUĞ	748	6	SÖYLÜ	601	108	LÜĞÜN	508	168	VRUPA	609	252
FOTOĞ	683	12	YOLCU	559	108	GÜRÜL	391	168	GÜNEŞ	583	252
OCUĞU	666	18	TOBÜS	548	108	RLÜĞÜ	386	168	OĞLUN	511	252
GÖZÜK	593	20	UŞUYO	380	108	OLDUĞ	7706	180	ŞUYOR	488	252
GÖRÜŞ	892	28	ÜŞÜNÜ	1264	112	BÜYÜK	2902	180	ONUŞT	468	252
GÖZÜN	472	28	ÜZÜNÜ	384	112	OCUKL	2023	180	LDUĞU	8212	270
ŞOFÖR	394	28	ÖRMÜŞ	374	112	DUĞUM	1672	180	YOKTU	1241	270
ÇOCUK	3085	30	CAĞIZ	443	120	APTIĞ	1405	180	CUKLU	753	270
CUMHU	612	36	LDÜĞÜ	421	120	UĞUNU	5192	189	KLUĞU	648	270
GÖTÜR	1125	42	ÜLKÜC	401	120	ÖĞRET	520	189	UKLUĞ	571	270
ÖRGÜT	802	42	LKÜCÜ	401	120	URUCU	468	189	ULDUĞ	506	270
GÖRÜY	540	42	DÜŞTÜ	385	120	UYGUN	420	189	YOKSU	487	270
GÖVDE	381	45	ÇIKIP	457	125	TUĞUN	402	189	OKUSU	409	270
GÖLGE	433	54	FÜSUN	1797	126	TURUC	400	189	OKUYU	374	270
ÜĞÜNÜ	914	56	GÜVEN	1027	126	ÜLMÜŞ	380	192	DÜŞÜN	4234	280
ÜŞÜNC	679	56	OYUNC	896	126	ĞIMIZ	1273	200	ÜŞÜND	1202	280
GÖRMÜ	446	56	BUGÜN	800	126	DOĞRU	2536	210	ŞÜNDÜ	1137	280
ÖZÜNÜ	441	56	HUZUR	764	126	PIYOR	987	210	ÜZÜND	1066	280
HÜZÜN	390	56	SORGU	465	126	GÜZEL	2753	216	ÖRDÜM	634	280
GÖREV	1101	63	SONUC	381	126	MUŞTU	1715	216	ÜRÜDÜ	516	280
GÖRDÜ	2408	70	DUYGU	1360	135	OLUYO	941	216	ÜDÜRÜ	398	280
ÖRDÜĞ	981	70	KUVVE	549	135	YLÜYO	600	216	OLMUŞ	1619	288
ĞUMUZ	568	72	YDUĞU	527	135	OLCUL	517	216	ÜLÜMS	741	288
OTOBÜ	548	72	UYDUĞ	416	135	LUĞUM	437	216	TIĞIM	1259	300
PTIĞI	1418	75	ÇÜNKÜ	1786	140	ÜŞÜNM	701	224	IŞIYO	731	300
DÜĞÜM	803	80	RDÜĞÜ	1261	140	TTIĞI	1068	225	ÇAKÇI	529	300
HÜKÜM	425	80	DÜĞÜN	1080	140	KÖPEK	737	225	DOKTO	515	300
GÖSTE	2246	81	DÖNÜŞ	610	140	CAĞIM	1511	240	IZLIĞ	449	300
UYGUS	633	81	NDÜĞÜ	517	140	HİÇBİ	2364	243	ZLIĞI	445	300
UVVET	543	81	ÜĞÜND	387	140	GEÇTİ	1125	243	HIZLI	403	300
YGUSU	539	81	BOĞAZ	635	144	HEPSİ	1114	243	DUĞUN	5346	315
OTOĞR	679	84	OĞLUM	396	144	YECEĞ	998	243	RDUĞU	911	315
ÖRÜYO	540	84	PACAĞ	379	144	SEVGİ	954	243	GÖNDE	814	315
ÖTÜRÜ	484	84	ŞTIĞI	910	150	BEHÇE	807	243	URDUĞ	788	315
GÖRÜL	414	84	IŞTIĞ	480	150	EHÇET	807	243	UĞUND	778	315
SOĞUK	581	90	CILIĞ	382	150	HEYEC	770	243	NDUĞU	512	315
MÜŞTÜ	807	96	BÖLGE	600	162	STİHB	749	243	UNDUĞ	487	315
GÜLÜM	750	96	ÖYLEC	579	162	TUTTU	532	243	CEĞİM	1469	324
ÖLÜMÜ	702	96	UYGUL	575	162	ZDIĞI	499	250	POLİS	1443	324
GÖRÜN	1592	98	ULUĞU	547	162	IŞIĞI	469	250	ŞÖYLE	982	324
ÖRÜNC	550	98	UTSUZ	500	162	BÜTÜN	4767	252	GEÇMİ	929	324
GÖRÜR	409	98	SUÇLU	443	162	ÜSTÜN	1587	252	BAHÇE	902	324
KÜÇÜK	1876	100	VVETL	437	162	GÖRME	1535	252	SÖZLE	630	324
GÖZLE	1904	108	UYUŞT	414	162	YORUZ	1468	252	SUZLU	556	324
UĞUMU	1167	108	YUŞTU	413	162	ONUŞU	744	252	BULUŞ	554	324
TUHAF	866	108	YÜZÜN	1827	168	ŞÜNCE	679	252	GEÇME	507	324
VİZYO	748	108	ÜNÜYO	755	168	YÜRÜY	676	252	MUTSU	488	324
CEVAP	681	108	OĞRAF	732	168	ÜYORU	651	252	USTAF	472	324
ÖYLÜY	644	108	ŞÜNÜY	612	168	UMHUR	613	252	ÜSTEŞ	472	324
HAFİF	605	108	ÜRÜYO	589	168	AVRUP	609	252	SAHİP	445	324

The bigram *gö* seems to be one of the most challenging n-grams. It consists of rare unigrams while it has a high frequency. If a bigram is seen more than one time in the ciphertext and its frequency is about the frequency of % 0.22 (100x25203/11M), it could be assumed to correspond to *gö*. Rougly, the bigram would be uncovered in the ciphertext of length 873 (2x11M/25203=872.9).

In the literature, the bigram *qu* is the most common sample way of attacking homophonic cipher for English [16]. Frequency of the bigram qu is % 0.2. However, the bigram would be expressed by 3 symbols. Therefore, it seems that precious information is obtained for beginning to attack.

The rest of the bigrams could contribute to solve ciphertext but their frequencies are too close. It seems better to turn back after trying to detect more symbols by using other n-grams.

Table 4.3 contains useful n-grams to solve ciphertext. Though the values are too close to each other, the trigrams *gör* and *uğu* could be evaluated as distinctive because of the frequency values.

Table 4.4 contains obtrusive values. The tetragram *cumh* would be expressed by 12 different symbols. However, detecting the tetragram would be easy. The beginning and ending letter of the tetragram would be replaced with only one symbol. Similarly, same rules are valid for the tetragrams *ptığ* and *vrup*. Moreover, the tetragrams *çocu* and *görü* have distinctive frequencies.

One of the most challenging n-gram seems to be a member of pentagrams. The pentagram *çocuğ* would be expressed by 6 different symbols. More interestingly, three letters of the tetragram (ç, c, ğ) would be repeated permanently in the ciphertext because each letter would be replaced with only one symbol. Detecting the rest of the letters (o, u) would be easier if the other letters are solved. Similarly, the pentagram *fotoğ* is a useful n-gram. The first letter and the last letter of the pentagram have a frequency of %1. Furthermore, the pentagrams *çocuk* and *gördü* have a distinctive frequency.

Another point that should not be ignored is that both the pentagrams *çocuğ* and *çocuk* consist of the distinctive tetragram *çocu*.

Distinctive n-grams exist as seen. It seems more meaningful to begin with looking for the bigram *gö*. Then, pentagrams and tetragrams should be attempted to detect. If pentagrams and tetragrams could be detected in the ciphertext, it provides significant advantage in the rest of the process. Even if these tetragrams and pentagrams do not appear in plaintext, distinctive bigrams and trigrams would most probably help to solve encrypted texts.

Suppose that the following ciphertext is given:

045 044 025 007 076 088 096 092 013 098 081 053 048 097 006 085 071 030 085 077 004 099 075 046
041 046 065 098 062 068 032 002 062 005 013 061 025 008 079 000 025 041 097 098 004 073 084 052
057 037 056 074 066 095 018 054 011 016 062 099 022 080 085 002 006 033 080 073 066 012 056 078
038 020 032 040 050 032 076 079 000 019 030 049 094 080 041 092 018 028 090 042 095 066 060 071
006 072 069 084 014 077 060 056 094 069 033 077 094 081 061 093 051 096 036 078 030 032 029 009
017 098 056 070 026 031 035 036 009 046 050 099 058 088 022 056 060 005 050 008 069 036 054 051
056 014 013 074 004 081 008 001 098 071 057 081 088 062 088 027 036 000 040 097 045 090 052 055
013 046 056 076 007 039 045 095 028 095 077 011 000 035 061 079 047 023 012 078 013 087 028 067
030 063 042 004 054 052 016 041 016 028 097 006 063 091 085 101 079 000 025 008 022 021 070 051
003 072 059 063 027 063 005 043 041 008 025 097 068 033 062 065 089 069 034 074 096 087 001 031
066 034 067 040 075 053 069 032 075 078 021 009 018 070 030 084 087 035 073 091 057 019 007 075
067 036 057 020 062 093 051 016 029 007 018 091 055 036 057 062 050 044 101 043 064 029 054 059
018 024 079 055 036 055 062 090 026 046 013 074 056 037 016 029 095 018 094 096 031 084 049 094
042 032 101 070 018 096 072 086 016 011 007 101 001 064 058 028 054 037 055 006 057 062 027 064
075 074 059 011 000 102 061 051 047 029 049 087 028 047 076 026 092 050 047 059 032 091 071 055
079 074 018 083 028 061 006 085 091 096 032 069 089 022 002 016 075 071 016 038 028 078 096 098
059 089 101 041 005 025 008 039 052 089 051 001 072 018 085 027 063 000 049 039 053 019 032 098
023 039 064 030 037 055 022 055 023 022 055 017 017 046 071 075 031 051 049 012 080 013 082 019
005 058 035 098 028 060 102 045 057 003 085 071 021 098 052 078 004 073 069 032 039 073 071 006
014 051 027 090 076 005 026 022 087 019 032 028 024 071 045 016 071 054 069 007 019 014 034 060
101 062 030 044 101 039 061 030 049 061 026 008 096 097 010 008 004 051 078 029 089 059 073 018
006 072 096 084 046 077 099 018 057 045 005 025 042 097 008 062 008 042 022 097 086 017 008 095
026 056 096 074 002 011 000 028 033 026 031 079 009 060 029 057 051 044 077 041 014 059 023 094
034 082 037 092 077 036 005 025 097 003 008 025 008 018 045 098 069 016 052 091 046 038 085 086

063 050 074 036 054 059 079 054 052 024 066 090 076 009 059 020 092 076 073 071 045 098 023 085
036 035 044 034 006 046 042 074 041 044 013 098 045 045 024 019 060 066 001 014 056 093 065 009
026 005 029 093 010 064 077 051 054 081 054 080 003 078 088 029 014 052 014 030 067 086 073 058
022 089 076 030 031 096 023 016 059 085 002 003 087 023 094 099 011 030 009 018 081 061 051 061
101 017 087 091 069 054 022 082 017 097 069 098 091 020 073 026 031 081 090 035 007 056 095 050
084 014 077 083 091 079 016 059 019 005 021 097 020 043 098 091 022 094 041 089 025 031 041 072
091 057 027 049 055 040 001 095 006 098 002 064 030 014 037 024 077 083 071 023 057 038 019 095
028 087 048 032 009 049 099 041 055 075 003 007 048 074 018 041 014 066 007 069 045 057 022 095
084 084 093 021 014 030 030 064 038 044 052 082 027 054 036 094 010 098 066 089 071 006 087 069
014 017 074 041 044 066 047 075 029 032 049 027 092 036 089 071 001 067 066 036 000 058 080 092
065 088 056 072 076 044 011 054 071 003 016 030 092 086 028 053 086 026 031 025 031 075 037 092
075 009 069 031 075 005 080 079 097 050 052 087 069 031 084 047 040 070 027 089 099 019 030 078
071 048 061 026 008 101 056 097 076 053 065 051 098 077 003 047 071 056 097 076 094 056 049 009
077 009 052 012 058 019 031 021 033 071 081 099 080 007 065 093 022 049 046 040 088 075 026 016
102 061 023 008 091 008 091 056 078 101 009 034 093 062 046 011 093 066 007 006 055 035 006 074
010 051 055 027 099 086 034 016 006 046 080 006 024 010 043 014 075 064 066 021 012 002 098 077
037 009 077 023 057 058 063 002 011 012 025 041 070 077 074 010 088 068 023 082 049 088 017 019
098 025 045 031 010 033 023 102 083 013 016 043 012 043 036 009 025 079 070 081 061 102 023 012
058 010 064 037 044 096 011 053 025 079 032 035 057 010 054 013 057 034 047 017 019 094 025 041
032 023 009 084 079 097 059 017 098 071 014 042 011 087 025 079 073 064 036 098 020 050 094 023
039 097 017 011 033 025 001 073 023 093 068 050 024 020 010 064 035 061 001 097 101 011 012 025
045 073 028 060 042 056 014 066 030 093 071 001 016 025 088 086 050 044 021 055 066 067 058 073
030 037 078 040 031 038 094 102 056 051 053 035 070 050 046 075 003 047 066 049 092 080 073 017
099 052 047 030 080 005 026 033 077 073 026 000 096 087 071 020 009 084 067 029 070 050 085 069
014 043 046 042 088 005 056 008 026 049 082 029 089 000 020 057 026 043 064 080 060 007 049 016
062 016 076 099 020 084 088 045 009 043 037 053 029 045 098 056 007 011 047 018 078 017 062 031
051 094 035 089 028 087 025 037 053 077 094 021 091 089 068 044 071 020 017 055 041 005 025 027
061 086 000 043 082 066 004 097 076 053 065 051 067 035 032 048 095 080 048 054 040 099 018 016
002 044 052 091 057 038 093 018 003 000 025 079 097 025 008 052 073 051 037 092 040 003 012 030
063 035 056 060 052 016 058 093 101 059 063 010 008 019 008 093 068 088 081 090 071 060 056 060
021 085 069 075 060 102 027 095 085 062 039 099 051 075 005 058 041 097 060 019 030 098 091 028
061 037 043 008 037 047 077 070 101 101 063 010 061 036 061 045 092 023 064 018 085 002 060 062
006 072 062 037 016 029 037 055 046 029 088 075 060 038 036 054 049 093 066 050 024 050 060 086
017 088 068 073 011 089 056 043 070 028 046 071 003 014 018 090 002 083 101 062 005 013 097 068
084 061 025 097 058 008 058 079 082 068 031 036 073 068 084 031 036 031 021 003 073 062 092 039
026 032 013 094 079 082 072 052 051 074 028 061 043 006 097 102 037 061 079 092 072 075 037 014
081 067 025 049 098 035 048 098 086 031 079 087 072 075 043 046 028 046 052 005 025 037 061 019
093 071 055 050 067 084 094 080 073 079 098 072 021 049 074 020 085 040 056 093 075 046 041 064
079 057 025 083 086 054 056 068 097 086 082 056 084 092 029 045 098 058 096 073 002 036 053 023

090 081 095 079 046 000 051 078 066 067 093 051 046 049 057 035 083 091 005 040 020 082 004 081
007 077 010 005 017 000 025 040 092 010 031 062 044 096 007 084 024 048 083 084 019 064 034 016
081 063 075 063 056 048 093 029 044 065 035 053 058 041 067 075 082 059 011 032 030 089 026 092
081 054 026 036 055 000 065 089 002 026 082 080 000 004 053 045 089 018 043 067 035 034 044 000
053 059 071 024 051 044 080 007 071 052 016 020 088 086 017 083 040 045 054 025 099 062 005 013
061 068 026 009 080 056 046 058 020 026 093 081 088 040 081 099 035 083 068 090 039 079 014 071
075 005 065 011 097 101 043 061 025 008 066 061 002 067 021 062 053 025 070 007 049 044 030 005
011 037 094 076 027 092 019 089 101 001 087 091 062 070 056 047 013 082 065 062 033 049 065 098
071 030 032 084 098 080 031 045 008 077 003 097 080 027 087 021 078 036 093 021 087 019 024 030
062 090 056 082 079 102 005 049 012 080 070 059 089 091 006 085 013 063 075 055 017 014 042 022
007 048 097 006 085 101 096 031 002 032 022 069 044 021 101 046 038 095 091 001 005 025 061 050
006 085 091 085 062 005 013 097 068 051 053 077 032 027 098 043 067 021 070 068 005 051 078 081
093 026 027 044 062 053 081 092 011 089 052 084 009 006 057 062 014 101 028 088 080 072 027 085
040 034 046 008 022 061 017 074 030 030 093 025 060 022 017 064 065 006 072 040 063 101 050 064
069 072 079 063 037 079 046 062 005 013 008 068 084 012 102 089 039 073 058 048 078 048 078 037
087 040 089 091 045 009 018 061 017 098 066 022 078 039 012 049 098 040 031 081 007 102 006 085
018 020 005 042 008 091 043 067 035 070 050 089 058 020 000 080 061 091 037 009 042 089 079 047
076 067 021 014 030 048 057 091 003 024 091 093 002 054 101 081 093 102 052 012 069 032 011 031
058 053 056 009 069 087 040 078 040 053 036 020 026 078 029 084 094 080 011 082 086 061 048 007
002 083 027 028 063 075 085 056 001 044 001 014 021 024 003 016 048 033 004 058 016 026 044 080
036 087 062 050 078 084 009 027 032 086 003 044 022 044 019 054 066 084 044 029 093 019 017 046
102 090 039 002 024 052 071 064 038 028 087 019 032 058 056 072 075 001 074 088 037 096 005 096
097 026 085 062 085 059 013 063 036 032 071 073 010 030 053 075 004 024 018 072 025 042 024 017
022 044 018 088 066 099 066 083 036 020 016 025 090 052 051 074 048 083 035 005 068 061 051 030
072 035 044 059 083 058 041 055 079 093 002 099 045 088 002 060 088 066 013 088 075 088 027 006
063 002 074 049 026 060 068 030 074 048 088 042 083 091 013 095 052 083 039 047 079 032 050 073
036 000 077 053 029 011 009 018 073 002 062 016 017 088 071 087 049 020 098 101 032 071 056 032
002 032 021 073 027 041 088 021 055 048 060 080 062 005 013 097 004 076 054 099 040 095 000 056
008 045 097 025 061 034 044 048 012 086 073 066 079 053 045 053 056 070 042 022 032 069 031 056
097 029 045 074 043 024 011 093 048 008 037 008 066 045 008 025 008 007 062 060 058 065 005 022
085 091 054 011 017 038 042 005 038 078 006 092 101 045 067 019 089 075 047 038 020 070 025 073
095 045 045 054 012 019 089 075 049 087 038 005 026 054 011 068 078 080 047 096 000 084 097 091
078 006 072 020 085 029 085 049 050 063 076 017 063 005 030 098 042 095 023 037 074 077 003 016
078 058 096 009 080 012 003 087 038 047 029 043 012 050 014 101 020 000 045 082 075 001 070 050
061 062 082 025 087 082 030 026 047 027 070 076 034 064 023 014 050 024 059 028 087 011 032 101
004 072 052 016 068 000 086 039 097 086 020 008 039 013 047 059 031 050 069 024 052 091 016 038
054 022 028 097 006 085 071 079 053 023 088 028 094 069 055 071 102 063 052 047 043 033 040 032
018 003 047 038 005 026 083 019 006 072 042 085 040 034 024 028 090 029 049 007 065 020 057 006
044 002 079 083 025 090 027 054 069 006 085 059 037 055 102 045 055 023 074 039 044 071 010 092

080 096 074 079 074 035 019 093 034 007 037 038 005 043 083 019 049 074 040 093 003 046 096 063
062 085 068 084 063 096 026 044 029 007 066 045 064 072 013 085 021 049 055 056 000 080 096 097
017 008 084 094 066 062 000 013 097 068 026 067 080 003 095 076 062 093 021 074 006 072 017 085
077 050 044 056 037 016 096 005 042 004 008 017 097 051 098 066 062 005 013 097 056 043 067 035
048 074 065 062 099 075 044 034 044 102 022 044 056 049 055 056 005 102 056 097 017 008 084 087
018 062 000 013 061 065 037 094 077 004 085 062 085 056 043 063 056 049 046 040 083 058 079 014
065 005 042 096 061 030 061 043 047 066 072 069 046 043 043 095 004 051 046 044 035 068 055 065
062 005 013 061 068 049 067 080 048 063 075 063 041 063 065 043 016 077 083 101 003 014 041 055
006 057 101 024 084 026 054 004 084 055 068 005 029 096 097 020 061 013 008 000 049 022 047 068
099 036 017 044 040 037 057 035 096 031 069 062 000 013 097 096 049 094 080 032 005 042 030 047
004 028 088 029 068 044 071 030 081 093 080 099 065 054 027 060 091 003 044 059 075 005 056 011
097 091 000 026 053 029 092 004 052 044 030 060 086 027 083 086 065 070 077 011 067 043 065 014
011 095 027 062 000 013 097 068 043 033 029 089 067 101 066 064 037 057 035 005 053 058 058 016
043 044 102 060 018 021 046 020 088 076 020 093 042 003 083 025 060 062 005 013 008 096 051 009
040 052 055 020 027 099 076 063 062 039 099 049 075 005 018 071 085 010 008 036 008 059 052 082
077 032 011 070 091 003 087 058 010 053 002 049 082 011 031 003 087 056 073 002 034 057 065 012
041 031 091 053 075 092 065 026 067 080 089 048 082 096 032 022 051 089 000 037 087 018 049 033
035 082 052 094 096 026 012 042 073 048 087 056 031 022 036 073 069 000 049 098 059 049 078 029
028 093 042 041 094 023 094 096 054 052 070 051 070 059 000 091 023 009 069 099 102 098 101 089
071 041 094 036 054 021 078 011 098 037 006 063 059 045 055 022 065 054 039 028 060 026 054 102
066 067 019 032 049 005 051 098 013 078 096 082 050 012 000 006 063 091 003 014 021 054 101 055
096 054 050 081 099 037 095 040 066 014 068 092 001 053 040 056 070 069 062 000 013 061 025 061
045 000 025 087 013 033 056 038 082 002 094 035 082 056 076 094 027 031 101 073 090 038 049 016
062 024 096 055 101 037 064 029 023 064 102 023 009 026 003 055 028 093 051 095 052 005 040 037
087 077 041 070 077 075 009 023 075 092 096 055 027 087 043 090 091 090 036 017 009 018 028 097
084 097 059 011 024 050 017 049 014 040 054 063 011 030 085 059 064 003 074 076 060 095 035 051
064 102 021 092 002 039 032 076 088 043 065 060 036 020 087 101 028 008 026 086 012 090 035 099
000 043 041 008 025 008 066 008 081 088 069 087 066 019 076 083 095 042 060 059 045 057 093 019
020 067 058 081 061 084 052 005 068 017 061 079 060 034 009 018 074 003 046 028 099 052 094 017
032 018 079 087 001 087 093 019 030 087 101 028 008 037 061 091 019 074 027 030 084 055 035 054
075 005 068 020 008 040 052 078 023 021 033 004 024 050 053 051 060 071 102 063 075 067 006 088
028 054 081 007 080 075 098 069 041 070 076 083 083 077 099 059 088 101 081 046 011 030 016 004
098 029 089 000 019 027 087 059 059 099 023 082 020 079 094 033 023 027 046 017 102 047 011 088
050 090 018 030 000 080 008 018 097 052 003 061 028 054 042 068 016 058 020 081 060 040 083 056
093 039 093 101 041 014 071 053 102 030 009 096 012 026 087 018 048 008 056 014 030 043 055 040
069 016 052 018 074 038 046 003 064 012 052 071 031 006 085 066 001 000 025 022 008 076 000 043
094 091 051 087 035 009 045 092 003 005 025 097 039 075 032 037 045 072 018 063 027 026 044 040
090 068 097 020 049 061 005 051 036 061 058 059 097 017 097 056 049 098 042 081 083 042 021 098
018 009 045 005 025 097 050 006 085 066 037 014 080 054 081 054 080 021 087 101 033

If the n-gram based attacking model is performed on the ciphertext above, the following consequences are obtained:

If the encrypted text is examined, the bigram *006 072* draws attention. It appears 7 times in text of length 3381. Moreover, the frequency of the bigram *006 072* (7x100/3381 = 0.2) is almost equal to the frequency of the bigram *gö*. Therefore, the bigram could most probably be the bigram *gö*.

Another pattern that comes one step forward in the ciphertext is constituted by the following trigrams:

062 005 013
062 000 013

The pattern appears 14 times in the ciphertext. Moreover, the trigrams specified above would only continue with the characters *097, 061* and *008* in the ciphertext. The pattern could only symbolize the tetragram *ÇOCU* because of the expression count.

The fact is that the tetragram *çocu* could only continue with the letters *k* and *ğ*. In the encrypted text, the tetragram *çocu* continues with the characters of *068, 004, 056, 065, 096, and 025.*

If the other trigram patterns that contain *025* are examined, the following bigrams are interesting patterns for an initial investigation:

008 025 008
061 025 061
008 025 061
061 025 008

It had already been detected that the characters *008* and *061* are used to encrypt the letter u. The pattern specified above would most probably be the trigram *uğu*.

Therefore, it could be said that the character *025* corresponds to the letter ğ and the characters *068, 004, 056, 065, 096* correspond to the letter k.

K O 080 K U
K O 042 K U
K O 102 K U
K O 029 K U

Detection of the letters (*g, ö, ç, o, c, u, ğ, k*) reveals some patterns specified above. The n-gram *KORKU* could only correspond to the patterns. Therefore, the characters *080, 042, 102, 029* could correspond to the letter *R*.

Thereafter, the following patterns are detected and solved respectively.

O K U 045 U Ğ U: OKUDUĞU
Ç O C U K 084 U Ğ U: ÇOCUKLUĞU
Ç O C U K L 012 R: ÇOCUKLAR
Ç O C U Ğ U D O Ğ 087 C 033 K: ÇOCUĞUDOĞACAK
Ç O C U K 043 A R: ÇOCUKLAR
Ç O C U K 049 067 R: ÇOCUKLAR
Ç O C U K 026 A R: ÇOCUKLAR
O K U L L 082 R: OKULLAR
Ç O C U K L 009 R: ÇOCUKLAR
K 063 Ç 085 K L 063 K: KÜÇÜKLÜK
Ç O C U K L A 035: ÇOCUKLAR
R Ü 002 G A R: RÜZGAR
O L 041 U Ğ U: OLDUĞU
Ç O C U K L 094 R: ÇOCUKLAR
A K 031 L L A R: AKILLAR
A Ğ L 098 R: AĞLAR
K O R K U 020 U C U: KORKUTUCU
Ü Ç Ü 059 C Ü: ÜÇÜNCÜ
Ö Ğ R 024: ÖĞRE
L A R D 092: LARDA

O L 053 R A K: OLARAK
T O R U 091 L A R: TORUNLAR
001 O Ğ U 050 G Ü N Ü: DOĞUMGÜNÜ
T O R U N 037 A R: TORUNLAR
K Ü Ç Ü K L Ü K L 046: KÜÇÜKLÜKLE
O L A N L 078 R: OLANLAR
028 U G Ü N: BUGÜN
D A L L A R D A K 007: DALLARDAKİ
G Ö Ç L 016 R: GÖÇLER
T O R U N L A R 070 M: TORUNLARIM
Ç O C U K L A 077: ÇOCUKLAR
Ç O C U K 051 A R: ÇOCUKLAR
088 L K O K U L: İLKOKUL
G Ü N L 055 R D 055: GÜNLERDE
K Ü Ç Ü K L Ü K L 044 R İ: KÜÇÜKLÜKLERİ
D E Ğ İ 076 İ K: DEĞİŞİK
E C 074 K L E R: ECEKLER
075 I L L A R C A: YILLARCA
Ç E Ş 099 T L İ: ÇEŞİTLİ
B E 052 O Ğ L U: BEYOĞLU
G Ü Z E L L 060 K: GÜZELLİK
G Ö 017 Ü R M E K L E: GÖTÜRMEKLE
30 Ü R K İ Y E D E: TÜRKİYEDE
039 U T L U L U K: MUTLULUK
K U Ş A K L A R 003 047 071 K U Ş A K L A R A: KUŞAKLARDANKUŞAKLARA
İ 019 T A N B U L L U L 047 R: İSTANBULLULAR
048 Ü Y Ü D Ü K L E R 083: BÜYÜDÜKLERİ
İ 036 T A 101 B U L: İSTANBUL
E 003 E B İ Y A T: EDEBİYAT
R Ü Y A G İ B 054: RÜYAGİBİ
D O Ğ 022 U Ş O L A N: DOĞMUŞOLAN
A C A 081 A: ACABA
S İ Y A S 057 T: SİYASET
B A Ş L A M A N 032 N: BAŞLAMANIN
D O Ğ U M Y 089 L D Ö 018 Ü 027 Ü: DOĞUMYILDÖNÜMÜ

095 L K K E Z: İLKKEZ
023 Ü S R E 034 G E R E D E: HÜSREVGEREDE
B A B 032 A L İ: BABIALİ
D O Ğ 079 U Ğ U: DOĞDUĞU
E R İ Y İ 038: ERİYİP
S 093 N E M A: SİNEMA
A N N E L E R İ N Y E T İ 086 T İ 040 D İ Ğ İ: ANNELERİNYETİŞTİRDİĞİ
T O R U N L A R I M 089 058 T O R U N L A R 089: TORUNLARIMINTORUNLARI
B U L U 066 D U Ğ U: BULUNDUĞU
K O M Ü N İ 011 T: KOMÜNİST
B Ü Y Ü D Ü K L E R İ N D 014: BÜYÜDÜKLERİNDE
R Ü Z G A R 073 N A: RÜZGARINA
B 090 R İ N İ N: BİRİNİN
069 E Y N E P: ZEYNEP
021 I L D Ö N Ü M Ü: YILDÖNÜMÜ
S E Ç M E N L 064 R: SEÇMENLER
K A L O R İ 010 064 R: KALORİFER

As it is seen, solving the encrypted text is possible on long enough ciphertexts. Hereby, all the characters of the encrypted text are solved and the following text is obtained:

DEĞİŞİKACABABUGÜNTÜRKİYEDEKAÇKIZÇOCUĞUDOĞDUAKILYELKENİ
NİSEÇİMRÜZGARINAKAPTIRMIŞDOSTLARDANBİRİNİNGÖZLERİKAZARAB
UİLKSATIRATAKILIRSAEMİNİMKİOMUZSİLKECEKBUDANEBİÇİMSORUDİY
ECEKŞİMDİBİRSORUDAHAACABATÜRKİYEDEBUGÜNÜNDOĞUMYILDÖNÜ
MÜOLDUĞUKAÇKIZVEKADINVARYAZIYAYANITLARINESİYASETÇİLERİN
NESEÇMENLERİNNEDESEÇİLECEKLERİNAKILLARININKÖŞESİNDENBİLEG
EÇMEYENSORULARLABAŞLAMANINNEDENİBUGÜNKIZIMZEYNEPBAKANI
NDOĞUMYILDÖNÜMÜOLMASIAHMETLEMEHMETTENYILLARCASONRABİ
RDEDÜNYAYAKIZIMINGELMİŞOLMASIBENDENİZİSEVİNÇTENMUTLULUK
UFUKLARININGÖKLERİNEDOĞRUUÇURMUŞTUİLKKEZSOBALIDAİRELERD
ENHAVALARSOĞUDUĞUNDAZEYNEPÜŞÜMESİNDİYENİŞANTAŞINDAHÜS
REVGEREDECADDESİNDEKİKALORİFERLİBİRDAİREYETAŞINMIŞTIKHENÜ
ZDAHAİSTANBULUNTANZİMATUZANTILIBİRİKİMLERİNDENSOYUTLANM

ADIĞIDÖNEMLERDİGAZETELERİNHEPSİBABIALİDEYDİBENDENİZDEMİLLİ
YETTEPEYAMİSAFANINGAZETEDENAYRILMASINDANSONRAKİKÖŞESİND
ETAŞBAŞLIĞIYLAYAZIYORDUMYAZILARIMIİSTANBULUNKUŞAKLARDAN
KUŞAKLARAYANSIYANBİRİKİMLERİYLERUHUNUNKANAVİÇESİNİGERGE
FLEMİŞVEGERGEFLEYENYAZARLARHENÜZSAĞDIREFİKHALİTSAĞDIFAH
RİCELALSAĞDIBURHANFELEKSAĞDIREFİCEVATSAĞDIHALDUNTANERSA
ĞDIESATMAHMUTSAĞDIHİKMETFERUDUNSAĞDIBİRKENTİNDEĞİŞMEYE
NANITLARIPARKLARIMEYDANLARITİYATROLARILOKANTALARIMÜZELE
RİOKULLARIOTELLERİİLEÇEŞİTLİDALLARDAKİSANATÇILARIBAĞLARAY
NIKENTTEDOĞMUŞOLANKUŞAKLARIBİRBİRİNEZEYNEPİNDOĞDUĞUYILL
ARDATÜRKİYENİNNÜFUSUİKİBİNİKİYÜZYİRMİÜÇMİLYONDUİSTANBULL
ULARINNÜFUSUDAHENÜZİÇGÖÇLERLEERİYİPSİLİNMEMİŞTİKISIKLIBEND
ENİZİNÇOCUKLUĞUNUNDAKISIKLISIYDIÇAMLICADAÖYLEBULGURLUDA
ÖYLEBAĞLARBAŞIDAÖYLEBEYOĞLUSİNEMALARIDAÖYLETÜRKİYEDEDE
ĞİŞİKKUŞAKLARDANKIZSAHİBİDEOLANAİLELERİNORTAKBİRFOTOĞRAFI
ÇEKİLEBİLSEVEBÜYÜKBİREKRANDAYANSITILABİLSEOKIZLAROKADINL
ARVEOANNELERİNYETİŞTİRDİĞİÇOCUKLARKENTLİBİRBİRİKİMDENYOKS
UNLUĞUNUZAYÇAĞIİLETOSLAŞMASINDANÇIKACAKÇALKANTILARIDUR
DURMAYASİYASETÇİKADROLARININGÜCÜYETERMİBUGÜNKIZIMZEYNE
PİNDOĞUMGÜNÜÇOCUKLARIMALAYIKOLABİLMEÇABASIYLAGEÇENBİR
ÖMÜRVEUMUTETTİĞİMTEKGÖRÜNMEZÖDÜLDEÇOCUKLARIMINBABALA
RINDANUTANMAMALARIBİRGÜNTORUNLARIMINTORUNLARIDAŞAYETB
ENDENİZİNBİRYAZISINAKAZARARASTLARLARSAŞUBİZİMBÜYÜKDEDEY
EDEBAKNELERSAÇMALAMIŞDEMESİNLERİSTERİMZEYNEPBASINKÖYDEİ
LKOKULÜÇÜNCÜSINIFTAYKENÖĞRETMENİNİNİSTEĞİYLEBİROKULTÖRE
NİNDEDİZİDİZİİNCİYİMGÜZELLİKTEBİRİNCİYİMADIMISORARSANIZÇETİN
ALTANINKIZIYIMDİYEBİRÇOCUKŞİİRİOKUDUĞUVEBAŞINDADAKIRMIZIK
URDELESİBULUNDUĞUİÇİNKOMÜNİSTPROPAGANDASIYAPTIĞIİDDİASIYL
APOLİSKARAKOLUNAGÖTÜRÜLMÜŞTÜOTARİHLERDEANKARADAPARLA
MENTODAYDIMUÇAĞAATLAMIŞVEHEMENBASINKÖYEKOŞMUŞTUMCANI
MZEYNEPİMBUGÜNDAHİBAZENRÜYALARINDAPOLİSGÖRÜRVEBİRLİKTE
GEZDİĞİMİZGÜNLERDEHEMENFARKEDERSİVİLPOLİSLERİDEKÜÇÜKLÜKL

ERİNDEÖCÜYLEKORKUTULANÇOCUKLARDİŞÇİYEGÖTÜRMEKLEKORKUT
ULANÇOCUKLARBEKÇİYEVERMEKLEKORKUTULANÇOCUKLARKÜÇÜKLÜ
KLERİNDEKORKUTULANÖZELLİKLEERKEKÇOCUKLARBÜYÜDÜKLERİND
EDEGENELLİKLEKORKUTUCUOLMAKİSTERLERKIZÇOCUKLARIORTAKBİR
KENTBİRİKİMİNDENYOKSUNOLARAKYETİŞMİŞKIRSALKESİMÇOCUKLARI
ANNELEROANNELERİNYETİŞTİRDİĞİÇOCUKLARYETMİŞÜÇMİLYONNÜFU
SUNYARISINDANFAZLASIDAKIZVEKADINAYAKLARIBAKIMLIOLANLARA
YAKLARIBAKIMSIZOLANLARBİRDAHAKİYILINONHAZİRANINDASİYASAL
GÜNDEMKİMBİLİRNASILOLACAKAMAOGÜNDEYİNEKİMBİLİRNEKADARK
IZÇOCUĞUDOĞACAKPAZARAKŞAMINIİPLEÇEKENLERHERHALDEBİLİYOR
LARDIRYAHYAKEMALİNİSTANBULUNSEMTLERİÜSTÜNEDEŞİİRLERYAZM
IŞİLKİSTANBULŞAİRİOLDUĞUNUBİZANSŞİİRİNDEİSTANBULYOKTUDİVAN
EDEBİYATINDADAİSTANBULUNSEMTLERİYOKTURYAHYAKEMALİNRÜY
AGİBİBİRYAZDIŞİİRİNİNBESTEKARIOSMANNİHATDAAHMETRASİMİNTOR
UNUYDUBİRKENTBİRİKİMİNDENARTAKALANBUKETLERZEYNEPEDEAYN
IGÜNDOĞMUŞOLANLARADADOĞUMYILDÖNÜMLERİKUTLUOLSUNNUTU
KLARBİRYANADOĞUMGÜNLERİBİRYANA

Homophonic cipher comes one step forward in the classical encryption methods because it generates the ciphertexts consisting of variable block sizes. This makes well known attacking models invalid. Although, the encryption method contains vulnerabilities for Turkish, it could clearly be said that the method is stronger than most of classical methods. Moreover, long ciphertexts are needed to cryptoanalysis. If long enough and uniform distributed ciphertext is given, distinctive n-grams would most probably contribute to detect vast majority of the letters of the alphabet. All in all, the method still maintains its resistance today against to frequency analysis attacks on short ciphertexts.

5. Conclusion

The key space of the substitution cipher is enormously larger than one of the most common modern encryption method DES. However, the block size of the encryption method remains stable and it is equal to the value of one. That would cause the defeat against to attacks based on statistical features of the source language.

Nevertheless, the most of modern cryptosystems such as *DES* and *AES* are insprired by the substitution cipher. Moreover, the substitution cipher constitutes the base for the substitution-permutation networks.

Herein, homophonic cipher is developed as an alternative to the substitution cipher method. Homophonic cipher extends the block size of the ciphertext in patches. Moreover, the block size is variable. This renders plaintext identification diffucult. Also, well known statistical features of the source language are invalid.

In this work, a novel attacking model for Homophonic cipher in Turkish is developed while main concepts of cryptology are demonstrated.

Herein, it is copied that Homophonic cipher has a key space wider than the most modern cryptosystems. All in all, without any hesitation it is figured out that the homophonic cipher still maintains its resistence today against to frequency analysis attacks on short ciphertexts. Long ciphertexts are needed to attack encrypted texts.

Meanwhile, the corpus size of the related work presented by Dalkılıç [1] is overperformed by this study. Therefore, more correct and more consistent results are obtained. As an additional deliverable, the results obtained towards the solutions of language based cryptographic problems also contribute to the linguistic studies.

References

[1] Dalkıllıç, M.E., Dalkılıç, G., "On the Cryptographic Patterns and Frequencies in Turkish Language", *Advances in Information Systems*, 144-153, (2002).

[2] Serengil, S.I., Akin, M., "Attacking Turkish Texts Encrypted by Homophonic Cipher", *Proceedings of the 10th WSEAS International Conference on Electronics, Hardware, Wireless and Optical Communications*, 123-126, (2011).

[3] Hoffstein, J., Pipher, J., Silverman, J., *An Introduction to Mathematical Cryptography*, Springer, New York, (2000).

[4] Menezes, A.J., Oorschot, P.C.V. et al, *Handbook of Applied Cryptography*, CRC, New York, (1997).

[5] Delf, H., Knebl, H., *Introduction to Cryptography*, Springer, New York, (2001).

[6] Stallings, W., *Cryptography and Network Security Principles and Practice*, Prentice Hall, New York, (2005).

[7] Acharya, B., Rath, G. S. et al, "Novel Methods of Generating Self-Invertible Matrix for Hill Cipher Algorithm", *International Journal of Security*, 1(1), 14-21, (2007).

[8] Eisenberg, M., "Hill Ciphers and Modular Linear Algebra", *Mimeographed Notes*, University of Massachusetts, 1-19, (1999).

[9] Acharya, B., Jena, D. et al, "Invertible, Involutory and Permutation Matrix Generation Methods for Hill Cipher System", *International Journal of Security*, 410-414, (2007).

[10] Dalkılıç, M. E., Gungor, C., "An Interactive Cryptanalysis Algorithm for the Vigenere Cipher", *Advances in Information Systems*, 341-351, (2000).

[11] Pommerening, K., "Kasiski's Test: Couldn't the Repetitions be by Accident", *Cryptologia*, 30(4), 346-352, (2006).

[12] Penzhorn, W., "A Fast Homophonic Coding Algorithm Based on Arithmetic Coding", *IEEE Transactions on Information Theory*, 45(6), 2083-2091, (1999).

[13] Penzhorn, W.T., Els, W.C., “A Universal Homophonic Coding Algorithm Based on Arithmetic Coding”, *Communications and Signal Processing*, 45(6), 2083-2091, (1994).

[14] Rocha, V., Massey, J., “On the Entropy Bound for Optimum Homophonic Substitution”, *Information Theory*, 93-94, (1997).

[15] Ryabko, B., Fionov, A., “Efficient Homophonic Coding*”*, 45(6), 2083-2091, *Information Theory*, (1999).

[16] Singh, S., *The Code Book: The Secret History of Codes and Code-breaking*, Anchor, New York, (1999).

[17] Günther, C., Boveri, A. et al, “A Universal Algorithm for Homophonic Coding”, *Advances in Cryptography*, 405-414, (1988).

[18] Jendal, H., Kuhn, Y., et al, “An Information-Theoretic Treatment of Homophonic Substitution”, *Advances in Cryptology*, 382-394, (1998).

[19] Sgarro, A., “Equivocations for Homophonic Ciphers”, *Advances in Cryptology*, 51-61, (1985).

[20] Hammer, C., “Second Order Homophonic Ciphers”, 12(1), *Cryptologia*, 11-20, (1988).

[21] Hammer, C., “Higher Order Homophonic Ciphers”, 5(4), *Cryptologia*, 231-242, (1981).

[22] King, J. C., Bahler, D. R., “A Framework for the Study of Homophonic Ciphers in Classical Encryption and Genetic Systems”, *Cryptologia*, 17(1), 45-54, (1993).

[23] King, J. C., Bahler, D. R., “An Algorithmic Solution of Sequential Homophonic Ciphers”, *Cryptologia*, 17(2), 148-165, (1993).

[24] Ravi, S., Knight, K., “Bayesian Inference for Zodiac and Other Homophonic Ciphers”, *Human Language Technologies*, (2011).

Appendix

In this work, the corpus of size 13.4 MB is used to obtain the Turkish n-gram frequencies. Also, the corpus consists of the sources specified below:

1. 120 articles of a columnist, *Çetin Altan*, from the Turkish daily newspaper Milliyet published between 20.01.2010 - 04.07.2010. (Available at www.milliyet.com.tr)
2. 37 novels of 9 authors as listed in Table A.1.

Table A.1 List of Novels Used in the Corpus

Index	**Author**	**Novel**	**Publishing Year**
1	Ahmet Altan	İçimizde Bir Yer	2004
2	Ahmet Altan	Kırar Göğsüne Bastırırken	2003
3	Ahmet Altan	Aldatmak	2002
4	Ahmet Altan	Kristal Denizaltı	2001
5	Ahmet Altan	Sudaki İz	1985
6	Aziz Nesin	Anıtı Dikilen Sinek	1982
7	Aziz Nesin	Borçlu Olduklarımız	1976
8	Aziz Nesin	Tatlı Betüş	1974
9	Aziz Nesin	Rıfat Bey Neden Kaşınıyor	1965
10	Aziz Nesin	Bay Düdük	1958
11	Aziz Nesin	Memleketin Birinde	1958
12	Aziz Nesin	Damda Deli Var	1956
13	Aziz Nesin	İstanbul'un Halleri	
14	Aziz Nesin	Sizin Memlekette Eşşek Yokmu	
15	Aziz Nesin	Gerçeğin Masalı	
16	Çetin Altan	Viski	1974
17	Gülse Birsel	Yolculuk Nereye Hemşerim	2005
18	Gülse Birsel	Hala Ciddiyim	2004
19	Gülse Birsel	Gayet Ciddiyim	2003
20	Orhan Kemal	Cemile	1952
21	Orhan Kemal	Baba Evi	1949
22	Orhan Pamuk	Masumiyet Müzesi	2008
23	Orhan Pamuk	Babamın Bavulu	2006
24	Orhan Pamuk	Hatıralar ve Şehir	2003
25	Orhan Pamuk	Kar	2002
26	Orhan Pamuk	Benim Adım Kırmızı	1998
27	Orhan Pamuk	Yeni Hayat	1994
28	Orhan Pamuk	Kara Kitap	1990
29	Orhan Pamuk	Beyaz Kale	1985
30	Orhan Pamuk	Sessiz Ev	1983
31	Rıfat Ilgaz	Hababam Sınıfı	1957
32	Soner Yalçın	Bay Pipo	1999
33	Soner Yalçın	Reis	1997
34	Soner Yalçın	Beco	1996
35	Soner Yalçın	Binbaşı Ersever'in İtirafları	1994
36	Yılmaz Erdoğan	Hijyenik Aşklar	2003
37	Yılmaz Erdoğan	Hüsünbaz Sevişmeler	2001

Biographical Sketch

Serengil was born in Istanbul on November 24th, 1986. In 2003, he graduated from Kadikoy Intas Lisesi in Istanbul. He began his undergraduate studies in 2004 at Istanbul Commerce University Computer Engineering Department. In 2009, he received his BSc degree in Computer Engineering from Istanbul Commerce University. He enrolled in MSc studies in Galatasaray University Computer Engineering Department same year. He is currently pursuing his MSc degree. Presently, he is working as a Software Developer at *Softtech*, which is a subsidiary compay of *Isbank Group* since August 2010. Also, he is the co-author of the paper entitled "*Attacking Turkish Texts encrypted by Homophonic Cipher*" which was published in the *Proceedings of the 10th WSEAS International Conference on Electronics, Hardware, Wireless and Optical Communications* held at Cambridge, UK in February 20-22, 2011.