Journal of Novel Applied Sciences
Available online at www.jnasci.org
©2013 JNAS Journal-2013-2-8/197-205
ISSN 2322-5149 ©2013 JNAS

A Preliminary Linguistic Analysis of Romanized Persian SMS Messages

Mohsen Akbari

Bakhtar Institute of Higher Education, Ilam, Iran

Corresponding author: Mohsen Akbari

ABSTRACT: SMS (Short Message Service) messages have become a common means of communication throughout the world. With new colors and tones in the communication world, Romanized Persian (RP) is an emerging code employed by Iranian texters. The current study is to illuminate the main findings of a linguistic analysis of RP SMS messages written by Iranian cell phone users. To this end, a corpus of 719 RP SMS messages was collected manually from informants. The data were transcribed and analyzed linguistically. Analysis was based on some linguistic features found frequently among the SMS messages and examples related to each linguistic feature were described in detail. The findings revealed that there are several main characteristics including punctuation, omissions, consonant writing, graphical means and symbols, contractions, letter repetitions, loanwords, and letter-number homophones. The implications of the study concern the way cell phone users in Iran use the language of SMS in their messages as well as the common linguistic features of RP SMS messages. The paper concludes by promoting the approach to analysis taken in this study as a basis for future research into SMS communication in Iran and even in global contexts.

***Keywords*:** Linguistic analysis, Romanized Persian, Short Message Service, SMS messages, Communication, Iran.

INTRODUCTION

Many individual users now have access to SMS messages as a written means of communication beside other forms of communication such as electronic mails, chats, forums and so on. The medium has become increasingly important for everyday communication within a large number of people in the world.

Many people have developed their own SMS message system both for formal and informal communications. The way language is used in SMS messages has raised questions regarding the genre, style, and linguistic features of the messages. Using language in computer-mediated discourses like chats, electronic mails and SMS texts significantly differs from the standard form of the language used. Among the main features of these texts composed of a non-standard use of language; are their brevity, limited length, written form, and structure (Choudhury et al., 2007).

SMS communication as an interpersonal communication between people who know each other might be influenced by several factors including brevity, the way a message is typed and written, the character limit of the messages themselves and the cumbersome text input. All these factors give the texts some specific features and make this terse and otherwise rude behavior acceptable. Today communicating by short messages over mobile phones is a common way to use language in an unusual form. Using a non-standard form of the language in these messages has led to disregarding grammar, punctuation and spelling rules. Cell phone users in order to type and communicate faster employ a large number of techniques to type the messages. Among the commonly-used techniques, abbreviations, shorter phonetic substitutions, deletion of words and characters, and using symbols, emoticons, and borrowings are some of the popular ways cell phone users employ as (Segerstad, 2004) mentions.

In addition, according to (Segerstad, 2004), the introduction and popularity of mobile phones and mobile text messaging has come to evoke hype about the impacts that the new technology is likely to have, just as with many earlier communication technologies. Central to the hype are concerns about the way that standard varieties and conventional linguistic and communicative practices are affected. Compared to standard written language, language in SMS messages is often reduced in a number of ways. These reductions often reflect a speech-like quality. In (Baron and Richard, 2003), among the electronically-exchanged written languages and discourse forms; messages sent over mobile phones via the so-called Short Messaging System (SMS) are of high significance today.

As (Fairon and Paumier, 2006) point out, the problem often raised by researchers concerning the study of these new forms of text is the shortage of reference corpora. This is especially true for SMS, for which the text collection is technically more complex than it is for emails, chats and forums. In fact, messages sent from phone to phone are difficult to collect, because it requires the collaboration of either the phone owners – but they are scattered - or the phone companies, but the latter of course have very strict legal regulations. As a consequence, there is no corpus big enough to permit large scale studies. Plus the need to an appropriate corpus, such writing venue has a cluster of basic characteristics. It is often composed quickly and in an informal style. Commonly, it evidences diminished attention to spelling, punctuation, and grammar. It involves exchanges between two interlocutors. Structurally, each exchange can begin with a salutation and end with a closing.It makes significant use of emoticons, abbreviations, and acronyms.

In all, communication by SMS messages has become increasingly commonplace and important in the world, yet the possible emergence of linguistic features in these means of written communication are neither well documented in linguistic-related research, nor widely covered in linguistic journals. Linguistic analysis of RP SMS messages widely used among Iranian cell phone users as a communication medium can open a new window to a new line of overlooked studies concerning linguistic analysis. The purpose of this study is to attempt a limited investigation into the linguistic features of real examples of SMS messages from Iranian cell phone users, which may yield some clues as to how people write such messages, and what are the common linguistic features of such messages. The aim of the current study is to analyze the characteristics of the SMS texts in particular the SMS texts in Iran. The scope of the current work is limited to the Persian SMS texts typed in a Romanized way. This study examines a corpus of data of real SMS messages as examples collected from Iranian cell phone users. The key text features of sources are analyzed separately under special headings to reveal any linguistic features commonly found in the SMS messages. The findings of the current study of SMS messages and more analysis in this regard will lead to establishing a new communication- and computer-related form of language in the world and especially in Iran. Short Message Service (SMS) is one of the most important ways of communication among many people in the world. SMS messaging has gained popularity among many people all over the world especially Iran in recent years. The various uses of SMS extend from keeping in contact with one another to communicating vital information to mobile customers, and sales reports and ads. Owing to the increasing number of SMS messaging in Iran, it is essential to investigate and find the ways Iranian cell phone users write the messages and analyze the texts regarding their linguistic features. Using the new SMS language that is popular with people who message each other using their phones - because this is easier than writing in normal language. RP SMS messages in Iran are in the form of phonetic writing, which is writing the Persian sounds using Latin letters.

No specific study has been done in this regard and there is lack of studies in linguistic analysis of RP messages. But to clarify what will go on in the present study, three related studies in line with the present study as the review of literature have been summarized here.

According to (Segerstad, 2004), results of an analysis of text messages written by Swedish users revealed some features that are characteristic to language in SMS, as well as various ways in which language is reduced are: shortenings, contractions and clippings, letter/number homophones, non-conventional spellings, accent stylizations, omission of punctuation and word spacing, excessive use of punctuation marks, emoticons, inflectional endings reduced. Creative, new abbreviations in analogy with unconventional abbreviations of English words were found, based on Swedish words. English words and phrases showed up in the middle of messages otherwise written in Swedish.

In another study, (Zelenkauskaite, 2004) conducted a research on the speech acts in relation to gender of users in SMS messages sent to a public Italian interactive television (iTV) channel. Written and publicly archived iTV SMS corpus available on the iTV channel's web page was investigated regarding their spoken properties, in particular speech acts, of the Italian language by gender. Moreover, a particular focus was attributed to gender differences in use of the language of iTV SMS. This study investigated differences in writing SMS on a public domain - interactive television (iTV) known as a phenomenon of convergence known as iTV SMS or iTV SMS chat. Italian iTV SMS data was evaluated. Preliminary results revealed gender differences in such a way that females

use more constatives, however male users use more directives. Employing a more detailed analysis, it appeared that women use more claims and male users used more direct and indirect requests. It appeared that male users used more greetings than female users did. These findings demonstrated that contrastingly what has been stated previously, there are gender differences in and it has been revealed in employing the analysis of the speech acts.

In a study by (Al-Tamimi and Gorgis, 2007), a counterpart of SMS messages, i.e. Latinized Arabic electronic messages were analyzed linguistically and just some general linguistic features were revealed as the main results. In this study, the focus was on the Latin script used in the messages and other general sociolinguistic features. What the present study is going to do is analyzing the messages linguistically in an atomistic way. I mean looking in detail and in depth at the RP messages and finding the most frequent and common linguistic features. At the outset, a definition for Romanized Persian (RP) should be formulated. Basically, it represents Persian language which over the last decade has been developing an electronically transmitted system of writing implementing characters, viz. Latin, which the English language uses. Such a code has been widely used by cell phone users, email senders and participants in chat rooms. The code, today, is a favorite linguistic choice for, especially, youngsters as long as there is no better competing code. Although Persian script has just started gaining ground electronically, yet preference is still given to the Romanized alternative in sending SMS messages. Unfortunately, there is no study tackling the issue of the language used in RP messages in Iran. The most important question raised in the study is:

What are the common linguistic features of RP SMS messages?

In an attempt to find answers for the above question, the present study will focus on the linguistic analysis of RP SMS messages. To do so, 719 RP SMS messages were collected from 10 cell phone users in Iran. They were analyzed linguistically and the most common linguistic features were described in detail using examples from the data corpus.

Definition of sms

SMS is one of the popular electronic communication technologies. According to the free encyclopedia of Wikipedia (2007), the Short Message Service (SMS), often called text messaging is a means of sending short messages to and from mobile phones. The term SMS is frequently used in a non-technical sense to refer to the text messages.

Limitations of the study

Among the main limitations of the present study, three were of utmost importance. The first limitation was a problem raised concerning the study of the new forms of electronic texts such as e-mails, chats, and forums; the shortage of reference corpora. This was especially true for RP SMS messages, for which text collection was technically more complex than it is for emails, chats and forums. In fact, messages sent from phone to phone are difficult to collect, because it requires the collaboration of either the phone owners - but they are scattered - or the phone companies, but the latter of course have very strict legal regulations. As a consequence, there is no corpus big enough to permit large scale studies. The next important limitation was restriction of these corpora to a certain and limited range of SMS users (they were all somehow English language M.A. students) which led to insufficient breadth and scope of the corpus of messages collected and some specific linguistic features in SMS messages. Furthermore, the final limitation was the stage in which the messages were collected, copied and transcribed manually. This introduced errors including typing mistakes or (in-) voluntary corrections. In all, the main limitations which were indispensable part of the study were the limited corpus, the restricted range of SMS users, and the way SMS messages were collected.

MATERIALS AND METHODS

The Present Study

The current study tried to do a linguistic analysis of RP SMS messages in Iran for the first time and fill the gap of related literature concerning Persian SMS messages particularly the Romanized ones. In this section; the data corpus, the methods that were used to collect a corpus of RP SMS messages and how they were analyzed linguistically are presented.

Data

The corpus of data was 719 SMS messages written in RP. These messages were Persian messages written with Persian words but in Latin letters and characters. They were collected from 15 Iranian cell phone users who

were M.A. students in English language. The RP SMS messages were the messages which were sent or received, then saved in the mobile phones of the users from April 2012 to March 2013.

Method of Data Collection

Due to the lack of a large public corpus of SMS messages as the main data source for use in the present study, the first task was to collect SMS messages that people had sent or received, then had saved in their cell phones. This was done so as to better reflect the language used in SMS messages and to increase the accuracy of analysis and design. The total number of SMS messages collected was 719 and the corpus is available in the appendix below. The corpus of SMS messages were collected from 15 Iranian cell phone user mainly M.A. students in English language. The SMS messages were elicited from 15 mobile phone users by sending a message which contained a request to send randomly 10-20 RP messages in the inboxes of their cell phones to the researcher's cell phone. The received messages were typed and transcribed manually in a Microsoft Word format for linguistic analysis.

Data Analysis

To obtain the linguistic features of the corpus of data, the 719 typed and transcribed RP SMS messages were analyzed and detected for common and main linguistic characteristics. All messages were examined for the most important features and examples related to each linguistic feature were described in detail. The findings of data analysis were summarized in the result section of the current study below.

RESULTS AND DISCUSSION

The following features were found according to the linguistic analysis of the RP SMS messages.

Punctuation mark:

Due to the nature of brisk communication, punctuation is extra difficult on a cell phone. It may or may not be known by the sender or receiver. Or may not be important to the sender or receiver, plus it takes extra time to use it and it does not significantly alter the message. But according to the data analysis in the present study among the 719 messages only 28 messages were lacking punctuation marks. The interesting point was using different kinds of punctuation marks or better to say excessive use of punctuation marks in RP SMS messages such as period (.), exclamation mark (!), comma (,), and question mark (?). This refers to not paying attention to the extra time to use punctuation marks by Iranian cell phone users and having interest to use marks in different ways while writing the messages. As examples are the two following messages;

(1)Salam. Tnx. Ta hadi. Saay mikonam ta 13om amadash konam. Dars ham moonde. Shoma che? Chizi shode?
(Hi. Thanks. To some extent. I am trying to prepare it until 13th. I have many studies to do. What about you? What's the matter?)

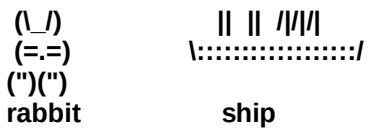
Slam shahram jon. nisti? Kojai? Aid jaii naraftin. Man alan khoonam. Asr mibinamet!
(*Hi dear Shahram. Where are you? Did you go somewhere on new year? I'm at home now. See you in the evening!)*

In these two messages, the cell phone users have made use of several periods and question marks while not using these marks would not alter the messages and would save the time in typing and sending the messages.

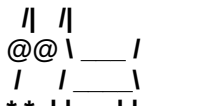
In addition to the common punctuation marks in most messages, some contained these punctuation marks in addition to mathematical symbols in a strange and unusual way. Punctuation marks and mathematical symbols were combined in some messages, and cell phone users instead of writing; had presented their messages in pictures or shapes made of the marks and symbols. This type of writing can be called *text art* because cell phone user have revealed their creativity and innovation in writing messages and created and coined new forms of messages.
For example;

(2)
><(((‘> **@**@++**@** **--<--@**
\\\\\\\\\/////////

fish **basket of flowers** **a branch of rose**



rabbit **ship**



dog

The above pictures are examples of using punctuation marks and mathematical symbols to create some shapes instead of writing messages.

Abbreviation (Contraction):

Some abbreviations and acronyms were used in RP SMS messages. Based on data analysis these abbreviations which were words or names made shorter by leaving out letters or using only the first letters of a word can be classified into two groups; first the Persian abbreviations, then the borrowed English ones.
As examples;
Persian abbreviations:
(3)
S *(Salam: Hi / Hello)*
Salm *(Salam: Hi / Hello)*
Fkr Knm *(Fekr Konam: I think)*
Chzy *(Chizy: Something)*

The data analysis revealed that the Iranian cell phone users rarely make use of Persian abbreviations and shortenings in their RP SMS messages. In contrast, they are interested in using Borrowed English abbreviations and shortenings. The analysis of data corpus in the present study showed that Iranian cell phone users as the participants in the current study tend to use English shortenings and abbreviations in their messages mainly due to their field of study; English language. So, it would be said that using English abbreviations and acronyms in the RP messages was as a result of cell phone users' degree and major.

Borrowed English abbreviations:

(4)
Tnx *(Thanks)*
Mrs *(a title for married women)*
Ms *(Miss)*
By *(Bye)*
U *(You)*
Dr *(Doctor)*
CD *(Compact Disk)*
C u *(See You)*
Ch4/5 *(Chapter 4 and 5)*
R u ok? *(Are You Ok?)*
Merc *(Merci)*

The analysis of data revealed that Iranian cell phone users especially the providers of the main source of data have tendency to shorten some English words and phrases and use them in the content of their RP SMS messages. The number of Borrowed English abbreviations was more than the Persian ones owing to the major of study of cell phone users in this study.

Smiley:

Using smiley or emoticon is an indispensable part of short messages including chats, e-mails, and in particular SMS messages. The cell phone users employ these shapes to show their feelings such as happiness, sadness, smiling, expressing love, giving thanks, kissing and so on instead of typing letters and using words. In the present study, only 4 messages contained smiley. It showed that Iranian cell phone users are interested in writing and using characters to show their feelings and emotions instead of smiley.
Examples of using smiley were shapes such as;

(5) :-) ; > :-* ;)

In addition to the smiley, other shapes were used to show the emotions and feelings of cell phone users. These shapes showed the creativity of users in coining new words, too. Although these shapes were classified under the punctuation marks section, they can be included as text art in which new shapes and pictures have been created and coined. The shapes of a heart and a rose using marks and symbols were interesting examples of such a creativity and coinage;

(6)

.*"".*"".
***Eide Shoma**
".Mobarak.*
"*.....*
*****

Ghalbetan labriz az eshgho shadio mohabat
(Merry new year)
(Hope your heart be full of love, happiness and kindness)

--<--@
Gole rozi baraye to ke khaili kaili dooooooooooset daram.
(A rose for whom I love very much)

Letter repetition:

One of the characteristics of the words used in the messages was repetition of some letters. This was done for different purposes especially showing feelings and emotions, interjection, attitudes such as delight, surprise, anger, shock, or disgust and onomatopoeia.
Some examples were;

(7)
Byyyyy *(Bye)*
Mamnoooonam *(Thanks a lot)*
Aziiiiiiiz *(Dear)*
Khoooobi *(Are you OK?)*
Jooonam *(Dear/ Darling)*
Dooooooooooset daram *(I love you)*

Although the number of such words and repetitions was not high, it showed that some Iranian cell phone users prefer to employ repetition of letters to show their emotions and feelings.

Spelling:

The way the RP SMS messages were written was not a kind of writing rather it was just writing the spelling of Persian words using English letters and characters. Using English letters instead of Persian sounds and letters did not follow a specific routine in the data corpus, rather for a sound or letter in Persian several English letters and sounds were used. In all, it should be concluded that Persian SMS messages are spelled and written in English letters and sounds, and the ways the Latin sounds and letters are used are personal.

So it can be said that Iranian cell phone users make use of English characters and sometimes strange spellings to write their messages in an English orthography. In another point of view, it can be claimed that the way

Iranian cell phone users use to write their messages is a type of phonetic substitution. So we can say that substituting phonemes in Persian SMS texts was among the other linguistic features. As examples:

(8)
Javab smset naresid. lotf kon ye tamas bgir
(I didn't receive you sms reply. Please call me)

Eshgh hamchon nabardist k b asani shoroo mishavad va b sakhti payan mipazirad.
(Love is a battle which starts easily and ends hardly)

Salam.merci.shoma khubi?karam tamum shode ta asr send mikonam.man shirazam az bacheha ziad khabar nadaram.
(Hi. Merci. Are you OK! I have finished my work I will send it until evening. I'm in Shiraz and I don't have any news about others a lot)

Loanword:

Due to the limited source of data and the specific cell phone users used in the present study (English language M.A. students), some loanwords and borrowed phrases were found in the corpus of data. These words and phrases were mostly borrowings from English language. For example;

(9)
Love Copy Abstract Sorry Good Luck
Mail Hi Class Proposal Testing
Site Form Send OK Syllabus
Computer Paper

Number:

Among the symbols used in the messages in this study was using numbers. Numbers were not only used as phone numbers, scores, and so on; but they were used as letter/number homophones. This type of letter/number homophone was a kind of phonetic substitution of some characters, letters, or words with English numbers. This type of writing can be called creativity in coining new words or as a text art. As examples:

(10)
2set daram
du3t daram
(dooset daram) :(I love you)
2shanbe *(doshanbe):(Monday)*
7te *(hafte): (week)*
2a (doa): (pray)
har 2*(har do): (both of them)*
Mer30 *(mersi) : (merci)*
13om *(sizdahom) : (13th)*
1i *(yeki) : (one of)*
3tare *(setare) : (star)*

According to the data analysis and example (10), Iranian cell phone users make use of numbers instead of letters, sounds and characters to show that they are faithful to the main feature of short messages that is brevity, or they want to show their coinage and fun in writing the messages.

Capitalization:

Putting the first letter of the first word into a capitalized form was the widespread feature of nearly all messages. Although in Persian writing there is no need to use capitalization, in the RP SMS messages due to the specific feature of cell phones when a cell phone user wants to write, it is written in capital letters automatically. In addition, owing to the automatic feature of cell phones if cell phone users write after some punctuation marks such as periods or exclamation and question marks with typing after a space the first letter of the words will be in

capitalized form. But if they do not use a space, the first letter of the words will not be in capitalized form. Examples related to this writing feature would help to clarify this feature.

(11)
CD haro dadam be hosayn.azashoon mitooni kopy begiri.bye
(I gave CDs to Hussein. You can take copies of them. Bye)

Salam. Kojai? Kay miay? Montazerim. By
(Hi. Where are you? When do you come? Waiting for you? Bye)

Mamnunammm. Arezooye khoobi ham mesl in baraye shoma aghaye akbari aziz. Omidvaram sale 87 sale shadmani, movafaghiat, salamati va shadkami baraye shoma bashad.
(Thanks a lot. The same best wish for you dear Mr. Akbari. I hope the year 87 be a year full of happiness, success, health, and felicity for you.)

Salam.khubid?are.hame na 8,7 nafar.karetuno dadam sare forsate behrtun migam goft chikar konim.
(Hi. Are you OK? Yes. Not all just 8, 7 people. I handed in your homework I will tell you in the right moment what did he asked for.)

Messages length:

One of the main features of cell phones is that the capacity of their creating messages section has a limited space. A cell phone user can only use 160 characters for one message. If people intend to send long messages, they are able to write them but in more than one page and by sending more than on message. In all, according to data analysis the RP SMS messages had various lengths. The longest message was composed of 463 characters and the shortest message was written in 3 characters.

Most messages were composed of 160 characters and only a few messages consisted of more than 160 characters. This showed that Iranian cell phone users tend to use fewer characters in their messages and follow the main feature of short messages, which is brevity. As examples, some of the shortest messages found in data analysis were;

(12)
mamnoooonam!
(Thanks a lot!)

doostat daram
(I love you)

Zang bezan
(Call me)

Kojae?
(Where are you?)

Mamnoon az lotfetoon!!!
(Thanks for your favor!!!)

Salam emailed ro lotf mikoni
(Hi Give me your mail please)

Che zibast b khatre t zistan. Love you!
(It is beautiful to live for you. Love you!)

Eyde shoma mobarak
(Merry new year)

Rooze zan mobarak!!!
(Happy women's day!!!)

Rooz ra khorshid misazad o roozegar ra ma.
(Sun makes the day and life makes us.)

CONCULSION

In this paper, a linguistic analysis of RP SMS messages in Iran was conducted. The novelty of the current work resides in analyzing the Persian SMS messages linguistically and reflecting the techniques used by Iranian cell phone users in the messages for the first time. Given the limited number of RP SMS messages, viz. 719, the study has shown that RP is an unstable code and hence given the name of a pidgin represented mainly by the Roman alphabet. The main common linguistic features found in this study were different ways to employ punctuation marks in the SMS messages, abbreviations, letter repetition, using numbers creatively, making use of symbols and smiley, capitalization, message length, the strange way of spelling, using loanwords. Although the present study was not a detailed and in depth linguistic analysis of RP messages and also it faced several limitations, it opened a window to the linguistic analysis of a newly-emerged code used in the electronic communication in Iran, which is a RP SMS message. Future studies are likely required to remove some limitations and to confirm the results of the current study.

REFERENCES

Al-Tamimi YA, Gorgis DT. 2007. Romanized Jordanian Arabic E- messages.The International Journal of Language, Society and Culture, 21. Retrieved January 3, 2012 from http:// www.educ.utas.edu.au/users/tle/JOURNAL.

Baron N, Richard L. 2003. IM and SMS: A linguistic comparison. Retrieved February 23, 2012 from http://www.ecommons.net/aoir/aoir.

Choudhury M, Saraf R, Jain V, Sarkar S and Basu A. 2007. Investigation and Modeling of the Structure of Texting Language. International Journal of Document Analysis and Recognition, 10: 3-4.

Fairon C, Paumier S. 2006. A translated corpus of 30,000 French SMS. In Proceedings of LREC 2006, Genova. Retrieved September 23, 2012 from http://www.uib.no/mailman/public/corpora.

Segerstad YH. 2004. The pragmatics of SMS. 9th International Pragmatics Conference, Italy. Retrieved December 20, 2012 from http://webhost.ua.ac.be/tisp/viewabstract.php.

Short Message Service. Retrieved March 7. 2012. from http://en.wikipedia.org/wiki/Short_message_service.

Zelenkauskaite A. 2004. Speech Analysis of Italian iTV SMS. Retrieved December 13, 2007 from www.indiana.edu/lingdept/conf.doc.