

## SMS Text Analysis: Language, Gender and Current Practices

Muhammad Shaban Rafi<sup>1</sup>

### Abstract

*This article tests the assumption that SMS language is like a pidgin in every speech community. The article also examines the assumption that a great motor of SMS lives among females whose lexical and morpho-syntactic choices are different from males. It further speculates influence of SMS language on language of media. One hundred messages were taken randomly from 20 cell phones and perceptions of 25 males and 25 females were recorded on an ordinal scale for analysis. The text was analyzed to look into lexicology, morphology and syntactic levels of texters, and influence of SMS on language of commercials. The results show that a novice intelligible language has evolved through SMS, which is influencing language of media. A significant difference is found between male and female texters' linguistic properties.*

### Introduction

Short Message Service (SMS) language tends to create a novice language, which has become an integral part of the multilingual world. It pursues simple sentences structure for communication. It is assumed that SMS syntactic and lexical choices by the texters are not so different from a child language. A child expresses his feelings through simple present progressive tense e.g. *mom eating* for 'Mom is eating' and *Eating* for 'I am eating'. The empirical data show that SMS language over-looks orthographic and syntactic rules of a language with a great emphasis on written sounds and compressions e.g. 8 for 'ate', 2 for 'to, two and too', 4 for 'four and for', *bcoz* for 'because' etc. The article also tests the assumption that advertising companies are pursuing SMS script to get across their product e.g. *Kahin na kahin tu hai...1 car 4 u* (Somewhere is one car for you), *Try Karna must hai* (You must try it.) etc. They go after simple sentences, written sound of words and bring code mixing in the advertising language.

### Aim of the Study

---

<sup>1</sup> Muhammad Shaban Rafi is a Research Associate cum lecturer and PhD applied linguistics research scholar at the department of English Language and Literature, University of Management & Technology Lahore.

Many researches on language and gender have been published but SMS text analysis in the domain of language and gender is open to investigate. The present study investigates lexical and morpho-syntactic choices of males and females SMS users. It looks into a) What types of words are used in text messaging? b) Whether or not SMS language follows mechanics e.g. punctuation and capitalization? c) Whether SMS language is more like speech or they are more rehearsed and ponderous qualities of writing? d) What is a relationship between lexical and morpho-syntactic choices of males and females texters? e) What is a relationship between SMS text style and language of media? Questions (a-c) are analyzed qualitatively and the following hypotheses are set, whereas questions (d and e) are analyzed quantitatively.

### **Research Hypotheses**

1. There is a relationship between lexical choices of males and females in cell phone text messaging.
2. There is a relationship between morpho-syntactic choices of males and females in cell phone text messaging.
3. There is a relationship between SMS and language of commercials.

Differentiations of linguistic choices between males and females have led to a number of language and gender studies in the recent years (Gray 2007; Montgomery 2000; Yule 2007; Kormos 2006; Trudgill 2007; Wardhaugh 2004; Macaro 2005; Napoli 1996; Doughty 2005; Ellis 2001; Jule 2005; Sudo 2007; Cameron 2005; Shehadeh 1999). According to Tanner (1990) and Wood (2001) women and men communicate very differently—in fact, it is sometimes stated that women and men communicate so differently from one another as if they were from different planets. Although at times differences in women's and men's communication styles seem to be constant and overwhelming, they are really quite minor. For example, both women and men can be nurturing, aggressive, task-focused, or sentimental. What is important to think about, however, is that women and men sometimes perceive the same messages to have different meanings. (Torppa, 2002). The question arises whether it is the difference of a perception or diverse linguistic choices that draws line between men and women communication approaches.

There are a number of studies examining the linguistic properties of e-mail and other computer mediated communication (Baron 2001; Yates 1996); however SMS language and gender have not received the same treatment. This article is an endeavor to contribute in this domain.

### **Over View of SMS Language**

Most text messages are not in the form of standard written discourse, but users are very effective in describing written sounds what they want their readers to perceive in their messages. Through the new written conventions of SMS, texters have developed a written form of sounds that replaces the ability to hear spoken utterances. The language used in text messaging has developed its own unique style as have email and chat-room languages e.g. *how r u? hop u'll b busy in ur stdy. Ma 2. When r v gona meet 4 dat work?*

### **Brevity of SMS Language**

The technical restrictions of text messaging have led to the development of language short forms in SMS communication e.g. limited space. Other reasons for short messages include the often difficult to manage interface and the fact that communications with close friends, partners and family members allows one to organize messages pragmatically as a common background exists. The messages serve to tie the group together through the development of a common history (Ling, 2000: 18). For these reasons SMS communication allows for a reasonable use of syntactic and lexical short forms, which save character space, or touches of the handset keys, as compared with using the full forms of words (Döring, 2002: 7). With this method of text production one saves time, money as well as effort and users are then more likely to use the service for subsequent messages. Döring also believes that abbreviations and acronyms fulfil a collective identity function whereby they require a special shared knowledge to be able to understand the language and consequently be able to use it. The adept use of these personalized language short forms is an indicator of group affiliation and a component of group identity. The language specific to SMS users often does not relate to standard language and the mass media thus label SMS communication as the secret code of the youth or as the big SMS action against long sentences (Döring 2002).

SMS allows users to abbreviate words without losing any meaning e.g. October and November can be shortened to 'Oct/Nov' etc. Similarly, punctuation, such as the full stop, is often unnecessary, as the end of a line will signify the end of an utterance. The resourceful use of punctuation as Koritti (1999: 15) describes it, 'constructing paralinguistic markers quite ingeniously as well as breaking orthographical conventions in an inventive manner appears to be a personal stylistic choice'.

### **SMS Compressions**

Written representations of the sounds and compressions are common phenomena in SMS language e.g. 'kt' instead of Katie. The choice to create a verbal illustration of this name tends to indicate that, at least in part, the user may be thinking of this utterance in its spoken form even though every other part of the text message may well be created within a written framework. One entirely 'spoken' aspect of text messages is not actually spoken at all, but rather emoted. To emote is to create a written representation of what the user is physically doing as he/she texts. Emoticons, such as :- ( , :- ) and ;- ) are a representation of body language, which would otherwise be missing from non face-to-face communication. These can change the meaning of a text message just as much as body language can change the meaning of verbal communication in spoken discourse. Texters may also take advantage of different phonetic spellings in order to create different types of verbal effects in their messages such as 'hehe' for laughter, or perhaps 'muaha' to express a frightening laughter. Letters and numbers are also often combined (or used alone) for compression and convenience e.g. 'See you' can be texted as 'CU ', 'Take can be texted 'TC' etc. According to Grinter and Eldridge (2001:17) if text messaging shares similar properties to e-mailing, we could expect these compressions to stabilize and become more widely known over a period of time.

### **Mode of SMS Language**

As with much online discourse, SMS retains both written and spoken language characteristics. As Rebecca Hughes (1996: 123) states, speech and writing usually take place in very different contexts. Written language is more prescribed than spoken language: authors have the chance to edit the words they write. SMS users make different word choices when writing than when speaking, as Biber (1998: 112) states, research indicates that 'all informational discourse has a high lexical variety in contrast to

interactive, affective types of discourse'. By studying language variations and uses, much of the conventions of text messaging language, a blurring of written and spoken discourse, can be better understood. SMS is unique with regards language selection, more like a written form of speech, as Biber explains, 'in terms of its linguistic characteristics, stereotypical speech is interactive, and dependant on shared space, time, and background knowledge; stereotypical writing has the opposite characteristics (1998: 25). Text messages however, tend to have more akin with the former. Koritti's (1999) work examines how 'reactive tokens' such as 'yeah I know', an acknowledgement that one is following what the other is saying, is very widespread in Internet Relay Chat (IRC) language. This feature of IRC is very common with SMS language. It shows how SMS language conventions resemble speech in many of its prominent characteristics. Like normal speech, text messages are structurally simple, fragmented, concrete, and conditional on situation-dependant reference. Many parts of spoken speech are eliminated through the process of abbreviation (Rintel and Pittam, 1997)

### **Method and Procedure**

The purpose of the study is to test SMS morpho-syntactic and lexical choices of males and females texters. It also investigates influence of SMS language on language of media. A sample of 50 subjects- 25 males and 25 females from ELT, linguistics and applied linguistics disciplines from a university was taken for the administration of a questionnaire. From male and female groups 100 SMS were taken for qualitative analysis of the data. The selection of the university and students was guided by convenience: a university where the students from all over Pakistan were studying was taken purposely. Factor analysis technique was applied to ensure the reliability of the instrument. 'Factor analysis aims to account for a large number of variables. It allows discovering the *factorial validity*<sup>2</sup> of the questions that make up each scale or construct' (Christine and John, 2004:416,417). Mean, standard deviation, standard error mean and independent sample t-test were calculated through Statistical Package for Social Sciences (SPSS).

Data on the target behavior of SMS texters were elicited and recorded through the administration of qualitative tasks and observation of the situations as conducted by

---

<sup>2</sup> The reliability of scales within questionnaires is measured by a correlation coefficient. Any coefficient  $\geq 0.70$  is strong, showing that these two scales can be considered reliable ones.

Bachman (1996), J.D.Brown (1996), Linn (1989), Popham (1981) and Selinger (1989). The items on the questionnaires were developed based on qualitative findings of this article. The students were asked to indicate their gender. Three value ordinal scales and yes or no criterion was adopted from Makino (1979:428) and Cazden (1986:227) over 15 questions, however 4 items were eliminated because they were reducing the reliability level of the instrument. Two questions (10) and (5) in the questionnaire were not treated in factor analysis because they were distracters (See appendix). Principal Component Analysis (PCA) was applied to transfer the original variables into a smaller set of uncorrelated components. The instrument was catalogued into 4 factors-morphology, syntax, lexicology and influence of SMS on language of commercials at 0.69 reliability level (See foot note). All these factors were considered to seek SMS users' perception on the questionnaire. The reliability of scoring procedure was adapted from Feldt (1989), Hambleton (1991), Orwin (1994), Shavelson (1991) and Traub (1994) in order to establish the extent to which score summaries represent systematic versus unknown or unintended sources of variability, by estimating classical and other sorts of reliability. According to Christine and John (2004:431) all the criteria must be used together to decide how many factors to keep for actual research. Therefore, correlation matrix, communalities, total variance, scree plot, component matrix, rotated component matrix, component transformational matrix and reliability scale were computed to reduce the data.

## **Results**

Qualitatively data collection shows that SMS language reflects abbreviations, capitalization and punctuation. Pronouns such as 'u' for /you/, 'v' for /we/, 'y' for /why/ and 'wen' for /when/ are frequently abbreviated. Conjunctions such as 'n' for /and/ and 'bcoz' for /because/ are also commonly shortened in SMS language. Similarly, at lexical level, the texters choice meaningful condensed forms e.g. 'intro' for /introduction/, 'bro' for /brother/, 'sis' for /sister/ etc. At syntactical level 'I am' for /am or m / is most recurrently used. Approximately 80% of the messages have no capitalization, another 12% had only first letter capitalization and the remaining 8% had complex capitalization. The SMS messages written by females are significantly more likely to have complex capitalization. Texters between the age of 20-25 years, are most likely to use capitalization in any form and also most likely to use first letter capitalization. They are

also the most likely to use punctuation in their SMS messages. Females use punctuation slightly more than male but the relationship does not appear to be significant.

The data show that there is a significant gender based difference in the number of words per SMS message and in the complexity of the messages. The data further show that females generally use lexically dense words in SMS messages. More than 74% of the messages were sent by males were simple one sentence or one-clause messages, whereas 51.61% messages were sent by females were consist of many clauses.

The data also analyzes that the texters often take advantage of the presence of both written and spoken aspects of SMS. The text messages reflect combine features of a written medium with features of a spoken medium e.g. *Hi dr xyz bro. How r u? M sorry 2 say that m sufring from high temp. So tis not possible 4 me 2 join u.* (Hi, Dr. XYZ brother. How are you? I am sorry to say that I am surfing from high temperature. So it is not possible for me to join you)

Quantitative data show that 38% males and 62% females prefer SMS for communication respectively as shown in Figure 1 below.

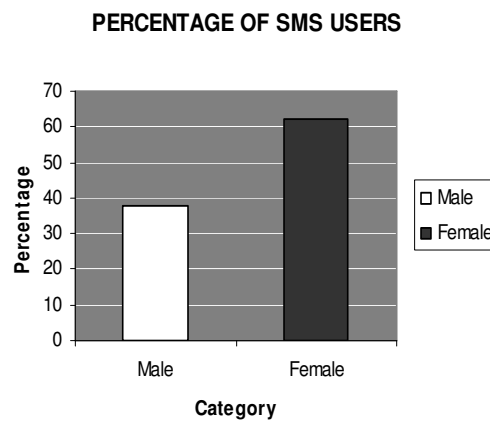


Figure 1

There is a mean score difference between males and females SMS users on the first three factors; that is to say, lexicology, morphology and syntax. On lexicology mean score of males is 2.04 with 3.85 Standard Deviation (SD); whereas mean score of females is 4.64 with 1.68 SD as shown in table 1 below. Small size of Standard Error Mean (SEM) 0.37 and 0.24 on lexicology between males and females respectively shows approximation of sample mean to population mean. Mean score of males on morphology is 1.84 with 2.35

SD and mean score of females is 3.80 with 0.42 SD; whereas SEM is 0.20 and 0.09 between males and females respectively. There is mean score difference on syntactic choices by males and females SMS users. Mean score of males on syntax is 4.92 with 5.72 SD and 0.19 SEM; however mean score of females is 7.24 with 1.24 SD and 0.07 SEM. Mean score of males is 1.64 with 2.85 SD and 0.09 SEM; whereas mean score of females is 1.56 with 2.69 SD and 0.10 SEM. There is very small mean score difference

Factor	Gender	N	Mean	Std. Deviation	Std Error Mean
Lexicology	1.00	25	2.04	3.85	0.37
	2.00		4.64	1.68	0.24
Morphology	1.00	25	1.84	2.35	0.20
	2.00		3.80	0.42	0.09
Syntax	1.00	25	4.92	5.72	0.19
	2.00		7.24	1.24	0.07
Ads_Language	1.00	25	1.64	2.85	0.09
	2.00		1.56	2.69	0.10

.08 between males and females perception about influence of SMS on language of commercials.

### Group Statistics

Table 1

The results show that 52% SMS lives among 15-19 years old population, whereas 40% SMS lives in 20-25 years old age group as shown in figure 2. It further shows that 4% and 2 % SMS lives among rest of the age groups.

**SMS USERS AMONG DIFFERENT AGE GROUPS**

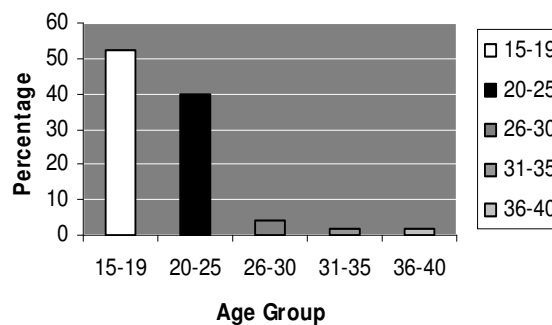


Figure 2



Table 2 below shows that two tailed independent sample t-test and significant values with 48 degree of freedom. The ‘t’ value on lexicology is 0.14 and 1.64 between males and females respectively. Similarly, the ‘t’ value on morphology is 3.19 and 1.12 between males and females respectively. The ‘t’ value on syntax is -1.20 and 2.90 between males and females respectively; however on the fourth factor, there is no difference of ‘t’ value between males and females, which is 0.57.

### Independent Sample Test

Factor		Levene’s Test for Equality of Variances		t-test for Equality of Means		
		F	Sig.	t	df	Sig. (2-tailed)
Lexicology	Equal variances assumed	1.24	.27	0.14	48	0.10
				1.64		0.94
Morphology	Equal variances assumed	.95	.34	3.19	48	0.89
				1.12		0.24
Syntax	Equal variances assumed Equal variance not assumed	.03	.87	-1.20	48	0.002
				2.90		0.19
Ads_Language	Equal variances assumed	1.14	.29	0.57	48	0.57
				0.57		0.57

Table 2

### Discussion

Figure 1 shows that a great motor of SMS lives among females. Ling (2000) concludes in his study that although men were early adopters of mobile telephones yet it is among the women that the great motor of SMS lives. Table 2 shows that there is significant difference between males and females texters on lexicology, morphology and syntax; whereas there is no significant difference in their perception about influence of SMS on language of commercials. Looking at particular age groups, it seems that 15 to 19 year old girls are particularly adept at writing complex and long SMS messages. Sattle (1985), Treichler and Kramarae (1983) and Rosenthal (1985) also concludes that females write longer messages and have a more complex structure than males. The analysis shows that there is a significant gender based difference in the number of words per SMS message and in the complexity of the messages. At the same time boys in this age group are particularly oriented toward simple “one thought” messages (Rich Ling, 2004).

Table 1 shows that there is mean score difference on lexical, morphological and syntactic choices between males and females SMS users; however there is unanimous opinion of the SMS users that mobile mediated language is influencing language of commercials. SMS language has given a birth to a novice language, which reflects people attitude towards their mother tongue and regional languages. Even the multinational companies have started advertising their products into SMS version and Roman Urdu.

### **Conclusion**

The article has analyzed that there is significant difference between males' and females' lexical and morpho-syntactical choices in cell phone messaging; however there is no significant difference between their perceptions about influence of SMS on language of commercials. Females are more skillful in writing complex, long and lexically dense messages than males. They have developed a unitary system of intelligible communication in the form of SMS language. It is also leaving backwash effects on standard English language and media language in Pakistan.

## References

- Bachman, L.F., Palmer, A.S. (1996) *Language Testing in Practice: Designing and Developing Useful Language Tests*. Oxford University Press.
- Baron, N. (2001) Why Email looks like speech: Proofreading, pedagogy, and public face. *Language, the media, and international communication*, 29.3. - 1.4. 2001, St. Catherine's College, Oxford, UK.
- Baron, N. (2000) *Alphabet to Email: How written English evolved and where it's heading*. London: Routledge.
- Baron, N. (1998) Letters by phone or speech by other means: the linguistics of email. *Language and Communication*, 18:133-170.
- Biber, D. (1998) *Variation across speech and writing*. Cambridge: Cambridge UP.
- Brown, J.D. (1996) *Testing in Language Programs*. Englewood Cliffs, NJ: Prentice-Hall.
- Cameron D. (2005) Language, Gender and Sexuality: Current Issues and New Directions. *Applied Linguistics Journal* 20\4: 482-502. Oxford University Press.
- Cazden, C. (1986) *Child language and education*. New York: Holt, Rinehart & Winston, 227.
- Dancey P. Christine and Reidy J. (2004) *Statistics without Maths for Psychology*. Using SPSS for Windows. Pearson Education Limited, 416-417.
- Döring, N. (2002) *Abbreviations and acronyms in SMS communication*. <http://www.nicola-doering.de/>.
- Doughty J.C and Long H.M. (2005) *The Handbook of Second Language Acquisition*. Gender- based Variation. Blackwell Publisher, 428-429.
- Ellis R. (2001) *The Study of Second Language Acquisition*. Oxford University Press, 202-204.
- Feldt, L.S. Brennan, R.L. (1989) Reliability in R.L. Linn (ed.), *Educational Measurement*. 3<sup>rd</sup> edition. New York: Macmillan, 46-105.
- Gray J. (2007). *Men are from Mars, Women are from Venus*. Thorsons Publishers, 5.
- Grinter, RE, and Eldridge, MA. (2001) *Proceedings of the Seventh European Conference on Computer: Bonn, Germany: Kluwer Academic Publishers*, 219-238.
- Hambleton, R.K., Swaminathan, H., Rogers, H.J. (1991) *Fundamentals of Item Response theory*. Newbury Park, CA: Sage.
- Hughes, Rebecca, (1996) *English in Speech and Writing: Investigating language and literature*. London: Routledge.
- Jule A. (2005) *Gender, Participation and Silence in the Language Classroom*. Oxford University Press, 125-128.
- Kormos J. (2006) *Speech Production and Second Language Acquisition*. Lawrence Erlbaum Associates, Inc., Publishers, 171.
- Kortti, H. (1999) *Internet Relay Chat and the Conventions of Spoken English*: <http://www.student oulu.fi/~hkortti/proseminar-final.html>.
- Ling, R and Helmersen, P. (2000) The adoption of mobile telephony among pre-adolescents and adolescents. *Conference on the social consequences of mobile telephony*, 16 June 2000, Oslo, Norway.

- Ling, R. (1998) *The use of traditional fixed and mobile telephony for social networking among Norwegian parents*. R&D report 33/98. Kjeller, Norway, Telenor.
- Linn, R.L. (1989) *Educational Measurement*. 3<sup>rd</sup> edition New York: American Council on Education and Macmillan.
- Macaro Ernesto. (2005) *Teaching and Learning a Second Language*. The Bath Press, 97-100.
- Makino, T. (1979) English morpheme acquisition order of Japanese secondary school students. *TESOL Quarterly*, 13, 428.
- Montgomery Martin, Duran Alan, Fabb Nigel et al. (2000) *Way of Reading, Language and Gender*. TJ International Ltd publisher, 77.
- Napoli Jo D. (1996) *Linguistics*. Oxford University Press, 196-198.
- Orwin, R.G. (1994) *Evaluating Coding Decisions*. The Handbook of Research Synthesis. New York: Russell Sage Foundation, 62-139.
- Popham, W.J. (1981) *Modern Educational Measurement*. Englewood Cliffs, NJ: Prentice-Hall.
- Rich Ling. (2004) *The Socio-linguistics of SMS: An analysis of SMS*. Use by a Random Sample of Norwegians. *Telenor R&D*.
- Rintel, E. Sean & Jeffery Pittman. (1997) Interaction Management on Internet Relay Chat. *Human Communication Research*: June97, Vol. 23 Issue 4, 507-535
- Rosenthal, C. (1985) Kinkeeping in the familial division of labor. *Journal of marriage and the family*, 47:965 -974.
- Sattle, J.W. (1985) *The Inexpressive Male: Tragedy or Sexual Politics*. Social Problems, 23:469 - 477.
- Selinger, H.W., Shohamy, E. (1989) *Second Language Research Methods*. Oxford University Press.
- Shavelson, R.J. Webb, N.M. (1991) *Generalizability Theory*. Newbury Park, CA; Sage.
- Shehadeh Ali. (1999) Gender Differences and Equal Opportunities in the ESL Classroom. *ELT Journal* Volume 53\4. Oxford University Press.
- Sudo Julia. (2007) Teaching New Tendencies in Gender Usage in Modern English. *ELT Journal* Volume 61\1. Oxford University Press.
- Tanner, D. (1990) *You just don't understand: Women and men in conversation*. New York: Ballantine Books.
- Torppa Cynthia Burggraf. (2002) *Gender Issues: Communication Differences in Interpersonal Relationship*. [www.hec.ohio-state.edu/famlife](http://www.hec.ohio-state.edu/famlife)
- Traub, R.E. (1994) *Reliability for the Social Sciences: Theory and Applications*. Thousands Oaks, CA, Sage.
- Treichler, P.A., and Kramarae, C. (1983) Women's talk in the ivory tower. *Communication quarterly* 31, 2:118- 132.
- Trudgill Peter. (2007) *Sociolinguistics: An Introduction to Language & Society*. Cox & Wyman Ltd Press, 78.

Wardhaugh R. (2004) *An Introduction to Sociolinguistics*. Blackwell Publishers, 313.

Wood, J. (2001) *Gendered lives: Communication, gender, and culture* (4th Edition). Belmont, CA: Wadsworth Publishing.

Yates, S. (1996) Oral and Written Aspects of Computer Conferencing in *Computer-Mediated Communication*, edited by Herring, S. Amsterdam: John Benjamins.

Yule G. (2007) *The Study of Language*. Cambridge University Press, 241-24.