

Use of Chinese Short Messages

Dafei Ma¹ Fumiko Ichikawa² Ying Liu³ Li Jiang¹

¹NO.10 (Jia), Datun Road, Chaoyang District, 100101 Beijing, China

²Arco Tower, Shimomeguro 1-8-1 Meguro-ku okia, TOKYO 153-0064, Tokyo, Japan

³Nokia House 1, No.11, He Ping Li Dong Jie, 100013 Beijing, China

motorfly5317@yahoo.com.cn, {fumiko.ichikawa, [ying.y.liu](mailto:ying.y.liu@nokia.com)}@nokia.com, jiangli@psych.ac.cn

Abstract. Short text message (SMS) as a key communication means in China received a lot of attention in research community. 114 subjects attended the study, sharing totally 10843 SMS they sent and received daily. We divided the SMS into two categories (instrumental and expressive), analyzed vocabularies, functions and effects of demographic factors and SMS categories on SMS lengths and found: 1) Top 400 Chinese characters occupied 85% and top 388 words occupied 73% in SMS. Punctuations appeared frequently (18%), while Smiley appeared very little (less than 0.1%). 2) People sent both instrumental and expressive messages regardless of their age. Female users tended to send longer SMS. Retired people sent longest SMS, followed by working people and students. People exchanged SMS with close friends and families. Expressive SMS have more words than instrumental SMS. 3) People over 40s exchange more SMS with children than with friends.

Keywords: SMS, China, instrumental, expressive

1 Introduction

Short messaging service (SMS) has seen a phenomenal growth in China after its proper release in 2000. In the same year, the total number of SMS sent by Chinese mobile users already reached up to 1 billion [1] and during the year of 2006, the amount reached 420 billion [2]. Since the penetration of mobile phone in China is still low and keeping increasing in near future, the SMS sent by Chinese users will keep growing too [1]. The primary reason for using SMS appears the same as in other countries: It is relatively cheap, convenient, and conspicuous [9, 10, 11].

Compared with communication behaviors of 137 million Chinese internet users [3], SMS was the very first technology enabling personal online communication in such a wide scale in China. The phenomena aroused many researches and discussions in terms of its impact to the society.

Just like many countries have reported, it is teenagers and young adults who drive the usage of short messages. In 2002, a local magazine Xin Zhou Kan sampled residents of large cities including Beijing, Shanghai and Guangzhou, and asked about their SMS use [4]. The results showed that two age groups, 15 to 19 year olds and 20 to 25 year olds had highest ratio of SMS users (83.3% of 15 to 19 year-old mentioned they

use SMS, while 89.0% of 20 to 25 year-old). In China, there is another factor which is the education. They found out that the higher the education of the user, the higher SMS usage rate. 31.1% of people with 9 years of compulsory education mentioned they use SMS, while it was 79.8% for the university students. Additionally, 49.3% of the non-SMS users mentioned that it is difficult to use. The research also investigated purposes of using SMS and the results indicated SMS are mainly used to greet people, chat, coordinate things e.g. daily work, and share contents e.g. jokes [4]. Lai and others studied college students in Shanghai, and reported their primary usage of SMS being making daily contacts, having a small chat, or showing care between acquaintances [5]. They also pointed out that SMS is preferred mode of communication as compared to other communication tools because SMS enables flexibility in terms of time of sending, time of reading, and whether to respond [5].

Another aspect which had received a lot of attention was the linguistic properties of text messages. In 2005, a collection of SMS poetry was published, followed by another collection of SMS novel [6]. On academic side researchers analyzed rhetoric, grammar and characteristics of SMS languages. In particular, seasonal greetings such as messages sent during Chinese New Year period had been analyzed by Zhang [7]. Messages during this festive time are peculiar, because of its content and the quantity. Fairly large amount of people would visit online to select short stories, jokes and poems developed by semi-professional SMS writers. Many sent messages in bulk as the year changes, and it was reported that over 10 billion SMS were sent on that specific moment [8]. Zhang classified these messages six categories by content and four by its format [7]. However, the study focused on this festive content, and had not extended to the content of everyday nature.

SMS is not always beneficial. Liu noted that although the peer-to-peer nature of SMS allows individuals to protect their privacy and enable personal discussions and free communications, he reminded that message can contain spam, and false content that would be harmful for the society [9]. Indeed, behind the impressive growth figure of SMS, spam messages have steadily increased and created social problems. Spam contents vary from harmless advertisement on flight tickets to false lottery winnings and in some rare cases, prostitution, but nevertheless, were out of control of individual's will to receive them.

There are a lot researches on SMS behavior in other countries. Ling mentioned SMS had found a niche in our communication needs, categories of the SMS can give us a clear picture of this niche [10]. Ling listed the most common themes used in SMS messages which are coordination, grooming, answers, questions, information, commands or requests, personal news and diverse other categories [10]. Grinter and Eldridge's research focused on teenagers and found their message categories are: communication coordination, planning activities, chatting and other (e.g., jokes, reminders, greetings etc) [11]. Those categories mentioned above give us some information about what people are talking about through SMS, however, they are somewhat not in the same level and some of them are overlapped.

The objectives of our study include 1) to identify the most frequently used words among Chinese mobile users; 2) explore characteristics of SMS Chinese users are sending and try to answer questions like: who are sending the longest messages and

what are the contexts or occasions to send them; and 3) what are the key purposes for using SMS in China.

2 Method

We used questionnaire to collect SMS and user profiles. Users can choose to use paper and pencil or computer to record their SMS and share with us. We instructed users to record SMS without any change but totally according to their original form. Users can decide freely SMS they would like to share with us, if they think some are too private to share, that is also ok. User profiles collected in the study include demographic information including age, gender, education background, occupation, location, and income and phone relevant information including phone model(s), payment for phone bill, operator, using frequency of applications, and main input method for Chinese SMS.

We don't use other efficient ways e.g. SMS gateway and forwarding SMS to collect SMS because of either higher cost or not willing to limit the study to certain group of users.

2.1 Participants

114 participants attended the study, sharing with us totally 10,843 SMS. Before recruiting participants, we pre-defined 5 users groups and size of each age group, which is according to the ratio of SMS users in mainland China. We used stratified and snowballing sampling method to recruit users. At the same time, occupations (students, working people and retired) were balanced in user groups. Table 1 shows a whole picture of the 114 participants.

Table 1. Demographic information of participants

Age	Occupation		
	Students	Working people	Retired
16~24(41)	31	10	0
25~34(40)	3	37	0
35~44(20)	0	20	0
45~(13)	0	4	9
total	34	71	9

2.2 Procedure

Firstly, we designed the questionnaires for profile and SMS collection. Secondly, we decided to adopt snowballing sample method, which can ensure the sample size, to recruit participants. Thirdly, we collected the data according to our plan. In the phrase, we at first contacted 10 coordinators who were typical representatives in the 5 age

groups and asked them to find more peers who would like to participate. The coordinators were also responsible for keeping contact with participants and ensuring the SMS recording and sharing. Totally, we found about 130 participants and finally 114 of them followed the study to the end. Numbers of SMS provided by each participant ranges from 50 to 200 items. A benefit of the snowballing method is that when we lose contacts with some participants, it is easy to reload with new ones. The data collection process lasted about two months. Finally, we transferred all SMS and user profiles to computer and analyzed them. In the analysis phrase, we focused on three topics: 1) vocabularies; 2) functions of SMS and 3) effects of factors like gender, age and functions on SMS length.

3 Findings and Discussions

In below, we presented our research results in four parts: 1) vocabularies; 2) categories; 3) effects of gender, age and category on SMS length; and 4) other results. In 1), we presented vocabulary results including the often used characters, words and punctuations in Chinese SMS. In 2), we described a method and result of categorizing Chinese SMS according to their functions. In 3), we analyzed the effects of gender, age and SMS category on SMS length and presented the results. In 4), we presented analysis result on relationships of SMS interlocutors across age groups. We also discussed the results after presenting them in each part.

3.1 Vocabularies

Table 2. Top 20 frequently used characters

Rank	Word	Meaning	Frequency	Rank	Word	Meaning	Frequency
1	我	I	5.5%	11	想	think	0.9%
2	你	You	5.1%	12	就	then	0.9%
3	了	(past tense)	3.6%	13	给	give	0.8%
4	的	of	2.6%	14	去	to go	0.8%
5	不	no	1.6%	15	没	didn't	0.7%
6	在	at	1.3%	16	呢	question ending	0.7%
7	吗	question ending	1.2%	17	说	to say	0.6%
8	吧	suggestion ending	1.0%	18	啊	ending	0.6%
9	好	good	0.9%	19	也	also	0.6%
10	是	be	0.9%	20	一	one	0.6%

Table 2 and Table 3 listed the most often used Chinese characters and 2-character words in Chinese SMS. Further analysis indicated that the top 400 Chinese characters

occupied 85% of the whole characters used in Chinese SMS and the top 388 words (including one-character words and two-character words) occupied 73% in Chinese SMS. The trend is very similar as in English [11].

Table 3. Top 16 frequently used 2-character words

Rank	Word	Meaning	Rank	Word	Meaning
23	什么	what	54	明天	tomorrow
32	现在	now	57	电话	call
37	知道	to know	62	你们	you
40	我们	we	63	可以	can
41	没有	don't have	70	就是	is
43	怎么	how	73	自己	myself
44	不是	be not	77	时候	when
47	今天	today	79	回来	to come back

Punctuations appear frequently (18%) in Chinese SMS and table 4 listed the most often used punctuations including comma, period, question mark, and exclamation mark. One thing to note is that Chinese users often mix using punctuations of half-width and full-width. In Table 4, the first comma, question mark and exclamation mark are in the form of full-width while the second ones are in half-width format. Surprisingly, English period is one of the most often used punctuation too. It might be because some phone platforms support only English period. Some users replace any punctuation with space in Chinese SMS, which is more convenient since space is always available and can be input directly in any input mode.

Table 4 also presented results about Smiley. Proportion of smiley was very small, less than 0.1% of entire corpus. Both vertical and horizontal Smiley were found, however, the result showed that in Chinese SMS, people primarily used vertical ones.

Table 4. Main punctuations and identifiable smiley

Rank	Word	Percentage	Rank	Word	Percentage
2	，	5%	320	:)	0.03%
3	。	4%	680	:~)	0.01%
7	，	2%	973	:{(0.01%
8	?	2%	2980	:~{(0.00%
9	.	2%	4509	:>	0.00%
11	?	1%	6973	:)	0.00%
13	!	1%	8740	:<	0.00%
18	!	1%	10666	:~)	0.00%
158	、	0	10658	^_^	0.00%

3.2 Categories

Based on the categorization of SMS in past researches [10, 11], we divided the SMS into seven categories according to their functions, purposes and moods. The seven categories are coordination, information, soft response, command and request, soft inquiry, personal information and emotional messages. Three researcher worked on the categorization of all collected SMS and everyone needed to categorize all SMS into the seven categories to make sure the categorization was valid. To keep coherence among researchers, researchers randomly selected some SMS to make trials. Everyone exercised on the SMS until the categorization coefficient reached 0.758 at the third round. During the categorization process, if two or all researchers put a SMS to a same category, we believed the SMS belonged to the category. And if the three researchers had different idea on a message, they would discuss on it until they agreed.

Based on natures of each category, we fatherly abstracted two general categories: instrumental and expressive SMS. The former five categories belong to instrumental SMS and the later two belong to expressive SMS in terms of their natures.

Table 5. Definition, example and proportion of 7 kinds of SMS

Category	Definition	Example	Percent age
1 Instrumental			
11 Coordination	Mutual planning of meetings & events.	<i>Laoguai, we will have a football match at 6 Am next morning at the playground. Join us on time!</i>	23.96%
12 Information	Ask or inform something that concerning both ends.	<i>Please tell me LaoLi's telephone number.</i>	25.10%
13 Soft response	Response to others' SMS.	<i>Ok. I think you might be busy, so I'll not disturb you any more.</i>	1.85%
14 Command and request	To ask for help directly.	<i>Buy me some bread!</i>	6.53%
15 Soft inquiry	Similar to Requests, but more descriptive and polite, and typically lengthy, describing situation and asking for some advice or help.	<i>To get an internship chance during vacation is really difficult. Could you give me a hand?</i>	0.28%
2 Expressive			
21 Personal information	To bring updates and in some cases gossips.	<i>Do you know XiaoLi has married?</i>	1.76%
22 Emotional messages	Chatting about personal or work lives, a way to show care to family or close friends.	<i>Forget the unhappiness, I wish you happy.</i>	40.53%

Table 5 listed all the categories with definition, example and proportions in the whole corpus. Emotional short messages occupied the highest proportion, followed by coordination and information. The other four categories are small in proportion and together they occupy about 10 percent. From the results, we can presume the key functions of SMS in China: 1) to show care to family and close friends and share personal and emotional issues; 2) to coordinate daily events and 3) to exchange and share information. When compared with SMS collected by Richard Ling [10], we can see both studies reported high proportion of coordination, but since the methodologies of SMS collection and classification are different, the proportion results are somehow different.

3.3 Effects of Gender, Age Group and Category on SMS Length

We also explored the effects of gender, age group and category on SMS length. Table 6 listed the average length of SMS of different user groups. Table 7 presented the MANOVA test results. The test results showed that:

1) *Female users write SMS with more words than male users.* As shown in table 6 and table 7, SMS written by female users are 1.26 words longer than those by male users, and the difference is significant according to ANOVA test (sig. =0.01). The result is compatible to our common sense, which is that female users are more sensitive and patient and inclined to write longer SMS.

2) *SMS sent by retired people have the most words, followed by working people, and SMS by students have the least words.* By MANOVA test, main effect of age is significant (sig. =0.00). Post-hoc test revealed that differences between any two groups are significant, which means retired people write SMS with the most words, followed by working people and the last is student group. The reason might be because the retired people have more urgent and strong needs to communicate with others on the fly. And since working people need to coordinate more daily events e.g. working issues than students, they might send more instrumental SMS than expressive ones, which will decrease their SMS length on average.

3) *Expressive SMS are longer than instrumental ones.* Averagely, expressive SMS have 16.92 words, significantly longer than instrumental ones as indicated by MANOVA test. The difference likely lies in the different functions of the two kinds of SMS. Instrumental messages are to coordinate daily issues and share information and a clear and concise style of writing would be applied. On the other hand, since expressive messages are to share emotional and personal issues, there would be more affection expressions involved in them, which would increase the SMS length. In the study, we also found that some people even tell a small story in SMS.

Interactions are also found as shown in Table 7. But they are hard to explain in the framework of the study. More controlled study might be needed to address them.

Table 6. SMS length by gender, age group and category

Variables		Average (SMS length)	S.D.
Gender	Male	14.82	0.46
	Female	16.08	0.20
Age Group	Students	14.28	0.17
	Working people	15.17	0.16
	Retired	17.00	1.08
Category	Instrumental	13.98	0.37
	Expressive	16.92	0.34

Table 7. Statistical test results

Source	Sum of Squares	df	Mean square	F	Sig.
gender	795.52	1	795.52	6.22	0.01
age	2,696.63	4	674.16	5.27	0.00
category	4,323.55	1	4323.55	33.79	0.00
gender * age	6,165.91	4	1541.48	12.05	0.00
gender * category	303.32	1	303.32	2.37	0.12
age * category	4,077.25	4	1019.31	7.97	0.00
gender * age * category	2,365.75	4	591.44	4.62	0.00
Error	1,371,513.54	10,720	127.94		
Total	3,794,732.00	10,740			
Corrected Total	1,417,788.78	10,739			

3.4 Other Results

Earlier we assumed that expressive messages are only for youth. However, the outcome showed that the expressive nature is found across generation.

We also looked at the relationship between interlocutors and Figure 1 showed the results. For users over 50's, children (most likely in 20s) became the primary SMS counterpart. It might be because SMS is a new technology introduced by some children to parents and parents don't necessarily to use SMS to communicate with their counterparts.

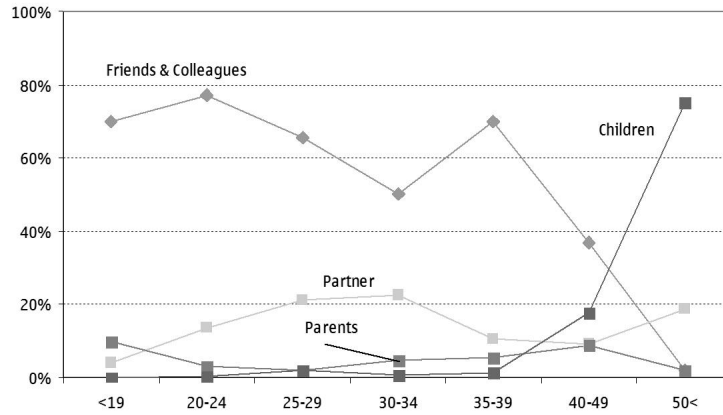


Fig. 1. Relationship between interlocutors via SMS by age

4 Conclusions and Future Works

In the study, we have collected 10843 Chinese SMS from 114 local users. We divided the SMS into 2 categories and 7 sub-categories according to their functions. Finally, we analyzed vocabularies, functions of Chinese SMS and the relationships between some demographic factors and characteristics of SMS. The results indicated: 1) Top 400 Chinese characters occupied 85% and top 388 words occupied 73% in SMS. Punctuations appeared frequently (18%), while Smiley appeared very little (less than 0.1%). 2) People sent both instrumental and expressive messages regardless of their age. Female users tended to send longer SMS. Retired people sent longest SMS, followed by working people and students. People exchange SMS with close friends and families. Expressive SMS have more words than instrumental SMS. 3) People over 40s exchange SMS exchanges more SMS with children than with friends.

In future work, we should conduct a field study of SMS users to understand their motivations and habits of using SMS. We believe if we can go to users and talk with them in their living or working contexts, we can have a better understanding of the data and results got in the study. Method of collecting SMS should be improved in the study. Since users can decide what to share and what not in the study, some bias might be involved in the study already. A better method to collect SMS might be for example asking users to share the last 10 SMS they exchanged on a day, which can improve research validity without increasing the cost.

Acknowledgments. We thank Nokia to sponsor the study. We thank Li Yongjuan from China Academy of Science for her help in the study process.

Reference

1. Zhou, Y., General Introduction of Chinese SMS Phenomena (in Chinese)
http://news.xinhuanet.com/newmedia/2005-05/12/content_2949398.htm, (2005)
2. Statistics of Chinese telecommunications (in Chinese)
http://www.mii.gov.cn/art/2007/01/22/art_27_28238.html (2007)
3. Annual Chinese Internet survey 2007 by CNNIC (in Chinese)
<http://www.cnnic.net.cn/uploadfiles/doc/2007/1/22/212245.doc>, (2007)
4. Did You Send SMS Today?
http://culture.163.com/partner/weekly/editor/020809/020809_64441.html, (2002)
5. Lai Y., Wang Y., and Wang S., Investigation of SMS Behavior of Shanghai University Students. *Journalists*, 2, (2004)
6. Ning, B., The First Novel about SMS Sells 5000 Yuan in China,
http://news.xinhuanet.com/mobile/2005-11/18/content_3799342.htm, (2005)
7. Zhang H., Discussion of the Spring Festival SMS, *Transaction of BeiHua University (Social Science)*, 2, (2005)
8. Jin J., Analysis of 10 billion SMS in 2005 Spring Festival: Information Supported Culture, *Shanghai Information*, 3, (2005)
9. Liu, J. How SMS Change Our Life. *Chinese Media Technology*, 2, (2005)
10. Ling, R.: The socio-linguistics of SMS: An analysis of SMS use by a random sample of Norwegians. In: *Mobile communications: Renegotiation of the social sphere*, London: Springer, (2005), 335-349
11. Rebecca, E. G., Margery, E.: Wan2tk?: Everyday Text Messaging. In: *Proceedings of ACM SIGCHI Conference on Human Factors in Computing Systems*, Florida, USA, (2003), 441-448