THE LENGTH OF TEXT MESSAGES AND USE OF PREDICTIVE TEXTING: WHO USES IT AND HOW MUCH DO THEY HAVE TO SAY?

Rich Ling

Introduction

Text messaging – or texting – via mobile telephones has become a fixture in many parts of the world. The ability to cheaply send text messages on a mobile asynchronous basis was adopted first by teens and is now spreading to other parts of the population. This said, texting is not an intuitive process. The interface is difficult to master, and the technology is being pressed into areas for which it was not necessarily intended. It is into this arena that systems of predictive texting have been introduced.

Predictive texting is intended to simplify text entry and to facilitate interaction. There are several issues which predictive texting potentially addresses. Most prominently, many writers – particularly those in the design tradition – have considered the contribution of predictive texting to the speed of text entry. Another issue is the length of text messages. Predictive texting might facilitate the amount of text that is produced. Research suggests that to some degree, predictive text messaging co-varies with longer text messages; however, this is a marginal contribution. Yet another issue is the socio-demographics and the technical sophistication of texters. Predictive texting has been adopted – and rejected – by different socio-demographic groups. The data show that younger and more advanced users of mobile communication make use of it. Finally predictive texting favors those who use standard forms of written language. All these issues will be addressed in this paper.

The standard layout on many mobile telephones is a 12 key pad (keys for the numbers 1 through 0 plus a “*” and a “#” key). To enter text requires that the user find a letter on one of the eight letter-bearing keys. If, for example, the user wants to enter “r”, he or she presses the “seven” key three times to cycle through the letters “p” and “q” before arriving at “r.” If we want to spell “hello”, we would need to key in the sequence 4 4 (h) 3 3 (e) 5 5 5 (l) 5 5 5 (l) 6 6 6 (o), a total of 13 keystrokes.
There are many alternatives to the standard “multi-tap” system. These include text recognition (such as used on a Palm), virtual and Bluetooth-connected QWERTY keyboards, thumb-pad QWERTY keyboards, voice recognition, and other forms of software input. There are even versions of the courtroom stenographer system for text entry that have been examined for use in the mobile texting world.

The most popular forms of text input on a standard 12 key mobile telephone are either the multi-tap system described above or a form of predictive texting. Data from Norway show that about half of mobile phone users use the multi-tap system. The others used predictive texting. The most popular of the predictive text entry systems is the commercial T9 system that consists of a dictionary calibrated to winnow down the key strokes into a single word, or perhaps a short list of words (Hård af Segerstad 2003). To use the example above, to spell “hello” using the T9 system, we first push the “four” key, which reduces the universe of possibilities to only those words starting with “g,” “h” or “i.” Next we push the “three” key, which further reduces the universe to only those words starting with “gd,” “ge,” “gf,” “hd,” “he” – the one we are after, “hf,” “id” – a possibility for the Freudians amongst us, “ie,” and “if.” Clearly, there are not many words starting with “gd,” “gf,” or “hd,” and so those paths are closed off.

There are, however, words like “get,” “geek,” “he,” “head,” “hello,” “id,” “if,” etc. that are still in play. As more keys are entered the list becomes shorter, and with luck, eventually there is only one word that is available. In the case where there are several options, the user can scroll down to the appropriate word, which is then inserted into the message.

There are, however, some quirks with the system. In the Norwegian version, if a texter wishes to use the word “you” – actually the highest-frequency word in text messages (Ling 2005), the first word to come up on the predictive software list is the Norwegian word for “it.” Some teens say that they do not bother to scroll down further, and instead of transmitting “Do you want to go to the movies?” simply send “Do it want to go to the movies?” Hård af Segerstad notes several problems with traditional predictive text systems in Swedish (2003). For example, predictive texting programs generally do not include dialect words or complex words that are not commonly used.
In theory, predictive texting should allow us to enter text quickly. Some models suggest that speeds can exceed 45 words per minute (Silfverberg et al. 2002). However, these models commonly do not take into account many real-life issues associated with text production, including differential manual dexterity, the needed learning curve, sophistication of the software, and time needed to figure out what we want to say, search out the proper keys, quibble with ourselves over composition, and clear up mistakes.

Comparing different types of predictive software and various gadgets for facilitating text entry, researchers report it is realistic to expect average users to achieve speeds of 5 to 10 words per minute (Butts and Cockburn 2001; Hård af Segerstad 2003; MacKenzie 2002; Pavlovych and Stuerzlinger 2004). Some expert texters, using different approaches, have been found to produce as many as 20 words per minute (James and Reischel 2001; Wigdor and Balakrishnan 2003; Pavlovych and Stuerzlinger 2004; MacKenzie et al. 2001). Research indicates that the traditional QWERTY keyboard – be it “on-screen” or a physical device – provides significantly faster text entry than does either letter recognition software (of the sort typically found on a Palm or Pocket PC) or text entry onto a 12 key mobile telephone (using either the multi-tap system or the T9 approach) (Cerney et al. 2004).

There has been a lot of discussion regarding how fast people can enter text. However, very few researchers have asked who uses predictive texting systems and the impact of predictive text entry on the length of messages and other characteristics of the communication (Ling 2005; Hård af Segerstaad 2005a; Hård af Segerstaad 2005b; Baron and Ling 2003). This paper focuses on these usage questions.

Our data come from a telephone survey of 1000 randomly-selected Norwegians aged 13 or older. The questionnaire gathered information on mobile phone ownership, usage patterns, and attitudes toward further technological developments in mobile telephony. One question on the survey asked respondents to provide the exact text of the last text message they had sent, including idiosyncratic spelling, abbreviations, and punctuation.
Another question asked if the respondents used predictive texting. A total of 417 messages were collected. (The remaining respondents had not sent a text message, were not willing to share the message with the interviewer, or had erased the message and thus it was not accessible.) The average message in the resulting texting corpus was 29 letters long (about 6 words).

The corpus of text messages was analyzed linguistically, including message length and use of emoticons, abbreviations, and dialectal forms. The primary hypothesis tested was that those using predictive texting would write longer messages. (Given the method of data collection, it was not possible to determine whether predictive texting took less time than traditional multi-tap input.) We were also able to examine predictive texting patterns in terms of the socio-demographic background of individual users and with respect to their attitude toward technology.

**The Role of Texting in Norway**

It is important to look at the use of text messaging within a particular social context, in this case, in Norway. Along with many European, Asian and Middle Eastern countries, Norway has made texting an important form of mediated interaction. Data show that text messaging is the most common form of mediated social interaction among Norwegians between the ages of 13 and 35. Almost 70% of teens between ages 16 and 19 say that they use texting on a daily basis. Only 30% of people in this age group say that they make daily use of the landline telephone. Where 60% of 20 – 24 year olds use texting daily in Norway, only 20% use the landline telephone each day. The picture changes for the elderly. For those over 67, about 30% use the landline telephone on a daily basis while only 10% report daily use of the mobile telephone for texting.

These numbers are even more remarkable when compared to texting behavior in the US. While mobile phone users in the US are at approximately the same levels as Norwegian users when considering mobile voice communication, less than a quarter the number of Americans report using texting on a daily basis (Traugott *et al.* 2006).
Data from Norway show that it is teens and young adults who are responsible for the bulk of these messages. Teens send a median of 5 – 6 messages per day. This is, however, a highly skewed distribution, with some individual teens sending as many as 50 or 60 messages daily. Thus, to provide a more realistic picture of the situation, the median score is more reliable here than the mean. While the numbers of adults using texting has grown in the last few years, the younger set still dominates the scene.

**Historical Background**

Texting on the 12 key mobile phone draws on a convention that was introduced into telephony as a mnemonic device in the 1920’s, when there was a shift from operator-assisted connections to direct dialing. Engineers at Bell Telephone, led by William G. Blauvelt, followed some independent telephone companies by introducing a system of words and numbers onto the phones themselves. Fearing that direct-dialing would lead to misdialing, the designers hoped to assist users who were at that point unaccustomed to dialing numbers (Farley 2005). A telephone number might be something like DRake-7923 (which, in terms of numbers, translates into “37 7923”). In the early days, the word – Drake in this case – often described the street location of the telephone exchange to which the telephone was connected. Eventually a number was added after the name, as the number of lines from each exchange grew. This system was reflected in popular culture, as in the song “PEnnsylvania 6-5000” by Glenn Miller and the Liz Taylor film “BUtterfield-8”.

This mnemonic approach only had room for 24 of the 26 letters in the English alphabet (three letters each for the “two” through “nine” positions). In the US system, “q” and “z” were dropped. This left the one position empty, with the “zero” reserved for direct access to the operator. In the UK, the “o” and the “q” were sometimes included at the “zero” position, and there was no “z”. In Australia, the “q” and the “z” were placed in the “one” position. In the US, the Bell system started to phase out the use of words in the late 1950’s, though as of 1977, almost 25% of the phone numbers were still using the name/number combination – speaking to the fondness people had for the names of their telephone numbers.
So long as letters were only a way to aid memory, they were little more than an interesting and perhaps endearing curiosity. With the development of interactive functionality in the mobile telephone world, the lettering took on new significance. However, the translation of numbers and characters to texting input was not clear-cut.

As we noted, the English alphabet has 26 letters, which fit three-to-a-key, spread over eight keys (with two exceptions). But what about upper and lower case letters? They necessitated a type of “shift” key – often the “*” key. In addition, languages other than English sometimes use additional letters. Thus, the three-letters-per-key idea needed to be expanded. In the work of the European Telecommunications Standards Institute (ETSI), the number “two” key hosts a large number of letters (along with the number “2”). These include: a, b, c, 2, à, á, â, å, æ, ç, č, ĉ, and three or four others not found in the standard Microsoft Word extended character set (ETSI 2003; von Numan 2003). In addition, there are complete character sets for Greek and Cyrillic, along with various punctuation marks and symbols for currency, mathematics, bracketing, and the like. In Europe alone, the keypad becomes complex, needing to accommodate 230 written scripts, most of which are variations on the Latin alphabet. This still leaves all of Asia, Africa, Oceania, and the Americas to deal with – all using just 12 keys. The authors of the ETSI standard note with a seeming sigh that while there are about 7000 languages worldwide, the telephone is lucky to deal successfully with 50 of these.

Results

In spite of the challenges of representing written language via a mobile phone keypad, texting is now a fixture in the daily lives of many people. Predictive texting is one design to simplify the input process. Drawing upon the Norwegian data described above, the remainder of this paper focuses on message length, looking at who uses predictive texting and to what effect. In addition, we consider some of the broader issues of texting in Norwegian society.

Message Length and the Socio-Demographics of Predictive Texting

The hypothesis we set out to test was whether use of predictive texting leads to longer text messages (compared with messages sent by those using multi-tap input). Figure 1
indicates difference in the median number of letters per text message for those people who use predictive texting and those who do not.¹

![Graph showing median number of letters per SMS for PT users and PT non-users.]

Figure 1 Median number of letters per message for users and non-users of predictive texting, Norway 2004

Overall, predictive texting users reported a median of 23 letters per message, while non-users reported a median of only 19 letters. While this difference falls just short of statistical significance (using the 0.05 level), it approaches that level.² Thus, the data suggest small differences may exist in message length when comparing people who use and who do not use predictive texting. Given the nature of the distribution, there is the broad hint that predictive texting leads to slightly longer messages though this finding is only on the outer edge of significance.

Interestingly, young adult females who do not use predictive texting created the longest messages. While the text messages reported by females aged 19 – 24 who used predictive texting had a median of 29 letters, those produced by the same age group who were not using predictive texting had a median of 46 letters per message. Males in

¹ Since the material on the number of letters resulted in a so-called poisson distribution, it is important to use non-parametric statistics, since they do not assume normality.

² This is a near-significant result when using the Mann-Whitney U test: Mann-Whitney U = 33727, sig. = 0.061.
the same age group showed similar trends: predictive text users reported messages with a median length of 18 letters vs. a median of 26 letters for those who were not using this system. In all other age groups, the ordering was either reversed or slightly higher for the non-users.

To summarize: Though we are not talking about statistically significant differences in message length, it is the young adults not using predictive texting who write the longest messages. Aside from this group, those using predictive texting generally write somewhat longer messages than non-users. The reason that young adults write longer messages, while not using predictive texting, may be that people who are proficient texters feel predictive texting actually slows their ability to input text. By analogy, think of a proficient typist who types half a word but then needs to choose from a list of alternatives before completing a word. Once a proficient texter is in the flow of texting, it is likely easier to simply concentrate on composition.

At face value, these findings appear somewhat confusing. Attempting to explain some of the patterns, we looked at use of dialectal (i.e., non-standard) forms in the text messages. People who used dialect forms in their text messages were significantly less likely to use predictive texting,\(^3\) perhaps because Norwegian predictive texting software is not designed to recognize dialectal variants of words. Young adults are the largest users of dialect formulations in their text messages. Many people in this age group have recently moved from the countryside (where dialectal forms of language are more likely to be used in speech) to the larger cities, either for education or jobs. However, they may retain communication links to their homes, where dialect is the norm. Retention of dialectal forms in their text messages may be one method of expressing their identity. Since young adults wishing to retain dialectal forms cannot rely on predictive texting to recognize their desired words (Hård af Segerstad 2003), it stands to reason that they are more likely to use the traditional multi-tap system for

\(^3\) Chi\(^2\) (1) = 4.346, sig. = 0.037. Those using predictive text were more likely to use emoticons than people who did not use predictive texting (Chi\(^2\) (1) = 5.812, sig. = 0.016). There were no significant differences between groups with regard to use of abbreviations.
creating their text messages. However, the dialect analysis shed no light on why the same group writes longer messages.

We return now to the broader question of who is using predictive texting. Comparing all males versus all females, there were no statistical differences. Approximately half of each group used predictive texting, while the other half did not. Turning from gender to age, the largest group of predictive text users was teens and young adults.\footnote{f = (1,951) = 232.3, sig. > 0.001.} Considering only people who had sent a text message in the last week, 96% of the 16-19 year old teen females used predictive texting while only 63% of the males did so.\footnote{Chi$^2$ (1) = 10.97, sig. = 0.001.} Indeed, this is the only statistically significant gender difference in the study. About 80% of both younger teens (13 – 15 years old) and young adults (20 – 24) used predictive texting. The use dropped off from there. Among the 55 – 67 year age group, only one in four reported using predictive texting. The levels were even lower among those over 67 years of age.

Predictive text users are also often active and sophisticated mobile telephone users. Figure 2 compares predictive texting use with use of voice and multimedia (MMS) functions on the mobile phone.
Active users of both voice and MMS functions were significantly more likely to use predictive texting than people who did not make active use of these functions. Moreover, people who sent larger numbers of text messages were also more likely to use predictive texting: a mean of slightly more than five text messages per day, compared with about two messages per day for those who did not use predictive texting.  

Those who reported using predictive texting were also more savvy about advanced telephone features than those who did not use predictive texting. Using a 5-point scale describing various tasks (such as the ability to set up a telephone to receive MMS messages or to read email on their mobile phones), users of predictive texting scored higher than non-users. Similarly, those who used predictive texting were significantly more likely than their counterparts (who did not) to be interested in using their phones for entertainment, user-driven directory assistance, information retrieval, and contact within their social networks.

---

6 Mann-Whitney U for voice = 89096.5 , sig. < 0.001, SMS = 78604.5 , sig. < 0.001 and MMS = 82227.5, sig. < 0.001.

7 On a five-point scale, the predictive texting users scored slightly more than three while the non-users scored only 2.36. (f = (1,951) = 139.4, sig. > 0.001).
The Gendering of Text Messages

Beyond the use or non-use of predictive texting, an important issue associated with the length of text messages is gender. Our data suggest that women generally write longer text messages than men.\(^8\) The median length reported by women was 25 letters, i.e., 7 letters longer than the median of 18 reported by males. Elsewhere (Ling 2004a, 2004b), I have suggested that gender roles are played out through the use of texting.

Data from several studies indicate that females – especially teenage and young adult females – are more expansive in their use of texting than males. Where men offer comments such as

> I think that there is something with SMS [= text messaging] . . . I can’t really do it. It is such short things (Bjørn, aged 40)

women do not generally have the same perspective. While women do write some short messages, overall their messages are longer than those of men. Here are some examples from the current corpus:

> super! Now we have landed at Steilende and the hot dogs are on the grill. The first landing from our own boat. M&MandT greetings. We are looking forward to saturday. :) (Female, aged 29)  


Hi! Are we still going to meet today? I don’t have more $ on my mobile after this msg. Just say when and where we should meet! (Female 19 years) Hey! Skal vi fortsatt møtes i dag? Har ik mer $ på mob etr denne mld! Bare si fra når og hvor når u vil møtes!

By contrast, male-authored messages are often more concise:

> Buy a harddisk (Male, aged 23) kjøp en harddisk

The pub doesn’t open today (Male, aged 32) Pubben åpner ikke idag.

\(^8\) Mann-Whitney U 27738.500, sig. < 0.001.
Admittedly, these examples from females are longer than the statistical median of 25 letters. However, in context of the statistical comparison between males and females, the examples clearly illustrate gender differences in orientation to the use of texting.

The other dimension affecting the length of text messages is the age of the sender, as show in Figure 3.

![Figure 3 Median number of letters per SMS (text) message by gender and age, Norway 2004](image)

Young adult women seem to be to the chattiest. Females under the age of 34 have the highest median number of words per text message. Women over age 35 use about 10 fewer letters per message than their younger counterparts. By contrast, males of all ages – aside from those over age 55 – are relatively stable at about 15 – 20 letters per message.

Our finding that women are more “verbose” in their texting language than men is consonant with conclusions other researchers have reached about gender differences in face-to-face spoken language (Treichler and Kramarae 1983; Sattel 1976; Clark 1995; Clark and Schaeffer 1981; Clark and Brennan 1991; Tannen 1991) and written communication (Biber and Finegan 1997; Argamon et al. 2003; Cheshire 2002; Coates
Gender differences have also been observed with respect to use of landline telephony. Women often talk on landline telephones longer than men (Claisse and Rowe 1987; Moyal 1989), particularly when the woman is the recipient of a call (Smoreda and Licoppe 2000). Research showing some of the same results has also extended to the use of electronically mediated interaction (Baron 2004; Herring 2003; Ling 2005). Moreover, in both spoken and written language, women are more likely than men to be supportive (Tannen 1991), use more forms of courtesy (Coates 1986), and be interpersonally “involved” with the interlocutor (Biber and Finegan 1997). It has been argued that female mastery of language provides them with a tool through which they can negotiate in various situations where they may otherwise be in a relatively powerless position (Cheshire 2002).

Across a variety of linguistic venues, there are gendered differences in the use of communication. Our findings regarding length of text messages are consonant with prior research. However, the introduction of predictive texting does not appear to have opened the floodgates of female texting nor has it provided males with a muse that would excite their passion for writing. It is simply a technique that is particularly popular among younger and more sophisticated users of mobile communication.

Discussion

The results indicate that predictive texting only slightly increases our chattiness. The development of predictive texting has been driven by the assumption that if we be simplifying input systems on mobile telephones, users will engage in either more or faster linguistic interaction. While our study did not engage with the speed issue, our findings provide some support for the hypothesis that predictive texting increases message length.

While this finding seems to play on traditional stereotypes, it is also true that women often have the responsibility for the organization of remote care-giving (calling the sick uncle or checking in on elderly parents) and for organizing various social events such as children’s birthday parties (Ling 1998). These conversations facilitate interaction, and they are often used in women’s networking that is, in some cases, down-graded and described as loose talk or gossip. Nonetheless, such talk can be seen as an indication of social integration as it is a part of the establishment of the standards of behavior (Gluckman 1963; Imray and Middleton 1983; Jones 1980, 194; Rakow 1992; Tannen 1991).
It may be the case that people have only so much to say. Looking at mobile communication through a somewhat broader lens, we can ask how much a technology facilitates or constrains the flow of communication. We might suggest, for example, that once a medium for communication has found its equilibrium, new fixes, gizmos and pricing structures will only change the flow of interaction in marginal ways. Developments in technical mediation only facilitate our ability to communicate; they do not give us the thoughts, ideas, or inspiration that is communicated.

Wide-scale adoption of the mobile telephone has allowed us to interact in places and at times that were previously impossible. Thus, the absolute flow of interpersonal communication has increased as a result of the technology. The threshold for communication is lower than with earlier technologies, and we may be communicating things that previously were left unsaid. This is not because these things were not relevant, but because before the adoption of the mobile telephone, in the time intervening between when the comment occurred to us and we saw (or could call) the individual to whom we wanted to talk, the item had either lost its relevance or simply slipped our mind. The immediacy of availability provided by the mobile telephone means that we now have what Licoppe (2004) calls connected presence.

Given a world in which mobile phone-based communication becomes the norm, advances in the technology – such as predictive texting – may further facilitate our use of mobile telephony. Alternatively, a leap in the net flow of communication may come by replacing today’s 12 key devices with portable or on-board keyboards (such as found on BlackBerries), which make lengthy textual input relatively easy. However, the lesson to be learned from the present study is that we must not be overly hasty in assuming that technological developments necessary lead to greater outpourings of language. Given a technical platform such as mobile communication, there are perhaps at best marginal gains or changes to be achieved through tweaking the technology.

We must also remember that communication events are limited by contextual constraints. There are situations in which people can sit and gab all day long, but these are special social contexts. Generally, the structure of our lives mean there is a limit to
the amount of time we can give over to a communication. A text message is squeezed into the time you are waiting for a ride and an email is dashed off before the meeting. (If you are a teenager, your ogre of a father may insist that two hours of talking on the phone is too much.) Given the other activities in our lives, our ability to dawdle over conversations are bounded. One of the effects of technologies such as texting – or email via mobile devices – is that we can squeeze social interactions into spaces where it was previously not possible to do so.

Our findings on predictive texting and message length suggest there may be limitations on the number of things that people have to communicate, at least via mobile phones. Predictive texting seems to lead to only slightly longer text messages. Using the mobile phone in a broad variety of settings means that its use has to fit into the opportunities that are available. The fact that texting is, by definition, a concise form of communication logically implies that facilitation of text entry (here, through predictive software) will only make marginal contributions to the production of text.

The notion that technology does not necessarily alter communication patterns can be further illustrated by a study my colleagues and I did in Norway a number of years ago (Ling and Hareland 1997). The goal of the project was to illustrate to customers how inexpensive it actually was to use landline telephones. In Norway all calls, including local calls, are metered. There is not the flat rate system presently used in the United States. Norwegian customers generally overestimated the cost of using the landline telephone. The experiment was designed to give users a true, real time sense of the cost of telephone calls they were making. If users feel constrained by what they wrongly consider to be a high price, they will curtail their use of the phone. This is a real world example of W.I. Thomas’ dictum, “If one believes something to be real, it is real in its consequences.” (Thomas 1931)

Users in the experimental group received a small metering device that displayed the cost of the calls they were making. The control group did not receive a metering display. The investigators determined the mean length of the calls both groups made. For the first month of the trial, the metering device had the expected effect. When
making local calls, those with the metering displays talked significantly longer than people in the control group. However, by the end of the experiment, the differences had disappeared. The people with the metering devices were back down to using approximately the same amount of time on the phone as the people in the control group.

It is difficult to make pronounced changes in communication patterns that are already entrenched. While customers in the study had a better understanding of the price issue by using the metering devices, over the long term, this understanding did not alter calling behavior. The landline telephony system has been in place long enough for people to have developed expectations with regard to how much and when it is appropriate to talk. Moreover, in our experiment, generally only the people making the calls (and paying the bills) knew their calls were less expensive than they had previously expected them to be. The people being called did not have this information. Even though the call recipients were not paying the bill, they continued to follow longstanding assumptions regarding how long a standard telephone call should be. Thus, entrenched telephone-related behaviors – on both sides of the conversation – led to relatively brief calls. As with predictive texting, technological aids to telephone conversations have the potential to alter consumer behavior, though in many instances, the changes are at best marginal.

Bibliography

Baron, N. and Ling, R. 2003. "IM and SMS: a linguistic comparison." in AOIR. Toronto: AOIR.


Licoppe, C. 2004. "'Connected presence: The emergence of a new repertoire for managing social relationships in a changing communications technoscope."


Ling, R. 1998. ""She calls, [but] it's for both of us you know": The use of traditional fixed and mobile telephony for social networking among Norwegian parents." Telenor R&D, Kjeller.

---. 2004a.""I have a free telephone so I don't bother to send SMS, I call:" The gendered use of SMS among adults in intact and divorced families." in Qualitative analysis of mobile communication, edited by Joachim Höflich. Universität Erfurt.


