

Running Head: TEXT MESSAGING AND IM

## **Text Messaging and IM: Linguistic Comparison of American College Data**

Rich Ling  
Telenor R&D  
University of Michigan

Naomi S. Baron  
American University

Please address correspondence to:

Naomi S. Baron  
Department of Language and Foreign Studies  
American University  
Washington, DC 20016-8045 USA

email: [nbaron@american.edu](mailto:nbaron@american.edu)  
phone: 202-885-2455  
fax: 301-229-4988

February 16, 2007

## **ABSTRACT**

While instant messaging (IM) via computers is well-entrenched in the United States, text messaging on mobile phones is a more recent technology in America. To investigate the emergence of American texting, this study compared text messages and IMs produced by American college students with respect to transmission length, emoticons and lexical shortenings, and sentential punctuation. We examine our findings in light of other statistical studies of texting and IM, and with respect to personal computer use in the USA that pre-dates text messaging.

**Keywords:** Instant messaging, IM, text messaging, texting, SMS, computer-mediated communication

**AUTHORS' NOTE:** The authors gratefully acknowledge valuable input from Howard Giles in preparing this Research Note. Correspondence should be addressed to Rich Ling, Telenor R&D, Snarøyveien 31, 1331 Fornebu, Norway; e-mail: [richard.ling@telenor.com](mailto:richard.ling@telenor.com) or Naomi S. Baron, Department of Language and Foreign Studies, Asbury 326, American University, Washington, DC, 20016-8045, USA; e-mail: [nbaron@american.edu](mailto:nbaron@american.edu).

Among teenagers and young adults, two popular forms of one-to-one electronically-mediated communication are instant messaging (IM), which is typically done via personal computers, and transmission of text messages on mobile phones. Text messages are generally limited to 160 characters, while IMs essentially have no upper limit. However, the input devices are not comparable. While texting uses one thumb (or two) on a small phone keypad, perhaps aided by predictive text software (Ling, 2005a), IM employs ten fingers on a full-sized computer keyboard.

Most text messages are composed all of a piece and sent as single transmissions. With IM, messages from a single interlocutor are commonly chunked into seriatim transmissions, yielding a sequence of IMs together constituting an utterance (Baron, 2004). For example,

isn't it nice [SEND]

to be in love [SEND]

in the spring [SEND]

With IM, once Internet access has been procured, messages do not have a per-unit price. By contrast, texting in the USA has been charged per message transmitted. Commercial plans now offer packages of messages or unlimited texting for a monthly fee, though texting charges remain in addition to monthly subscription costs for voice calls. Outside the USA, both voice calls and texts are unit-priced.

Text messaging began in Europe in 1993, with commercialization of the GSM mobile phone network. Popularity of texting in Europe and Asia has been especially high among teenagers and young adults: texting has been cheaper than voice calls, and many young people lacked ready access to IM programs on personal computers. As of 2005, the USA had 76.2 PCs

per 100 inhabitants, while Europe averaged 30.2.<sup>1</sup> Internet usage shows a similar divergence: 69.6% of the American population versus 38.6% of Europeans.<sup>2</sup>

In Europe and Asia, mobile phones are well-established. For example, 95% of Norwegians own mobiles, including 100% of teenagers. Approximately 70% of Norwegians aged 19-24 report daily use of text messaging (Ling & Haddon, in press). In the USA, as of 2005, 69% owned mobile phones. Approximately 4% of everyone in the USA and 18% of those aged 18-24 used texting daily (Traugott, Joo, Ling, & Qian, 2006).

IM was introduced in the 1980s at several American universities. However, the technology was not popularized until development of PC versions: ICQ in 1996 and America Online's own program in 1997 (Baron, 2003). AOL's free software, AIM, is the most common platform among American teenagers and college students, though alternatives include MSN Messenger and Yahoo! Messenger. By 2004, 48% of American youth aged 12 to 17 used IM daily (Lenhart, Madden, & Hitlin, 2005).

IM is gaining ground in Europe,<sup>3</sup> but the balance still heavily favors texting. By contrast, IM is ubiquitous on American college campuses. Mobile phones are also becoming de rigueur, and most students have at least experimented with text messaging. However, as of late 2005, mobiles were used on US campuses more than twice as much for voice functions as for texting (Baron & Ling, in preparation).

Linguistic analyses of texting have appeared for several languages (e.g., German: Döring, 2002; Swedish: Hård af Segerstad, 2002; Norwegian: Ling, 2005b, British English: Thurlow & Brown, 2003). Among the stylistic features noted are abbreviations, acronyms, emoticons, misspellings, and omission of vowels, subject pronouns, and punctuation. Since texting in the USA is comparatively new, collecting texting data in the early 2000s was problematic. By

contrast, IM has been amenable to research (e.g., Baron, 2004 and in Canada, Tagliamonte & Denis, in preparation). These two statistically-based IM studies, along with Thurlow and Brown's texting analysis, reported that abbreviations, acronyms, and emoticons were less prevalent in young people's computer-mediated communication than suggested by the popular press. To move beyond media hyperbole, we need corpus-based analyses of such features as abbreviations and punctuation. By collecting data from similar populations, we can compare the linguistics of texting and IM.

A second lens through which to view texting and IM is prior familiarity with computer technology. In the USA, middle-school and especially high-school students have been encouraged for over a decade to produce written assignments on computers. College freshmen arrive as proficient typists. Baron (2004) argued that the written sophistication of students' IM is consonant with prior experience on the same computer keyboards used for school compositions. The question now is whether American texting is colored by prior experience with IM which, in turn, reflects proficiency with word-processing.

Our three research questions were

*Research Question 1: What are the linguistic characteristics of text messaging in the USA?*

*Research Question 2: How do these characteristics compare with IM in the USA?*

*Research Question 3: Is extensive prior experience with word processing and IM reflected in American texting?*

Findings reported here involve three linguistic areas.<sup>4</sup> The first is length: How many words and characters were there per transmission, how many one-word transmissions, and how many sentences per transmission? The second is emoticons and lexical shortenings: How often were emoticons, acronyms, abbreviations, and contractions used? With contractions, how many

contained apostrophes? The third is sentential punctuation: How much punctuation appeared at the ends of sentences, and was it haphazard or principled?

A convenience sample of text messaging data was collected in Fall 2005 using paper diaries distributed to undergraduates at a large public mid-western university. Students were asked to record exactly all text messages they sent over a 24-hour period. Twenty-five completed diaries were returned (22 females, 3 males). Only the female data were analyzed. The resulting corpus contained 191 text transmissions, with 1473 words. IM data were drawn from a corpus collected in Spring 2003 from undergraduates at a mid-sized private university on the east coast. From the full corpus, a random female sample of 191 IM transmissions was extracted, containing 1146 words.

Text messages averaged 7.7 words, while IM transmissions averaged 6.0 ( $F_{1,385} = 10.97$ ,  $p = 0.001$ ). Recall that for IM transmissions, sequencing of consecutive transmissions is common. Therefore, while individual text messages were longer than individual IMs, average length of a complete conversational turn (combining seriatim transmissions) was longer in IM.

The mean number of characters per transmission in the texting data was also significantly larger than in IM. Text messages averaged almost 35 characters,<sup>5</sup> while the IM mean was just under 29 characters ( $F_{1,404} = 5.80$ ,  $p = 0.02$ ). Another factor contributing to message length is one-word transmissions. There were significantly fewer one-word text messages (7 out of 191, or 3.7%) than one-word IMs (36 out of 191, or 18.8%) ( $\chi^2(1) = 21.83$ ,  $p < 0.001$ ).

Almost 60% of texting transmissions contained multiple sentences, compared with 34% of the IM transmissions ( $\chi^2(1) = 22.29$ ,  $p < 0.001$ ).<sup>6</sup> The mean number of sentences per text-message was 1.76, while the mean for IM was 1.27 ( $F_{1,380} = 38.62$ ,  $p < 0.001$ ). Since IMs are commonly sent as consecutive transmissions without added cost, this finding is not surprising.

Of the 1473 words in the texting corpus, only 2 were emoticons (both smileys). In the IM corpus, out of 1146 words, there were 5 emoticons: 4 smileys and 1 frowny face. In the texting corpus, only 8 acronyms appeared: 5 instances of *lol* (“laughing out loud”) and 1 each of *ttyl* (“talk to you later”), *omg* (“oh my god”), and *wtf* (“what the [expletive]”). The IM corpus had 4 acronyms: 3 cases of *lol* and 1 *ttyl*.

There were marked differences between samples in use of abbreviations. The IM corpus had no clear-cut abbreviations specific to online communication. One case of *b/c* (“because”) occurred, but this abbreviation is commonly found in informal writing, and predates computer-mediated communication. Three miscellaneous lexical shortenings appeared: *ya* (“you”), *prob.* (“probably”), and *em* (“them”). However, these forms frequently appear in the informal speech of American college students, and are not specific to IM.

By comparison, the texting corpus had 47 unambiguous abbreviations: 26 instances of *U* (“you”), 9 cases of *R* (“are”), 4 examples of *k* (“OK”), 6 occurrences of *2* (“to” – both as a word and as part of “today”), and 2 instances of *4* (“for” – as a word and as part of “before”). In addition, the texting data contained words with vowel deletions: 2 instances of *b* (“be”), and 1 case each of *latr* (“later”) and *ovr* (“over”). We cannot determine whether these examples represent intentional shortenings or input errors. Nearly a dozen texting examples involved miscellaneous lexical shortenings, e.g., *Sun* (“Sunday”) and *tomm* (“tomorrow”), only some of which reflect casual speech, e.g., *ya* (“you”).

Contractions (e.g., *can't* instead of *cannot*) typically appear in informal speech and writing, and are shorter to type than full forms, especially when omitting the apostrophe. In computer-based IM, apostrophes require only a single keystroke, while needing four key taps on mobile phones. We calculated percent of full and contracted forms against total potential

contractions. For apostrophes, we scored only use in contractions, not possessives (e.g., *Mildred's*). In texting, 84.7% of all potential contractions were contracted. In IM, only 68.1% were contracted ( $\chi^2(1) 9.246, p = 0.002$ ). Far fewer apostrophes appeared in texting contractions than in IM: 31.9% versus 93.9% ( $\chi^2(1) 47.784, p < 0.001$ ).<sup>7</sup>

We examined punctuation at the ends of transmissions and the ends of sentences (since many transmissions contained multiple sentences). We also tallied use of question marks at the ends of semantically-interrogative sentences in comparison with use of periods, exclamation marks, or equivalent punctuation (ellipses, dashes, commas, and emoticons) at the ends of declaratives, imperatives, or exclamations.

Texting and IM followed similar patterns, with the proportion of texting punctuation always lower than in IM. Total sentence-final punctuation was 39% for texting and 45% for IM. Transmission-final punctuation appeared in only 29% of text messages and 35% of IMs. However, for transmissions containing multiple sentences, the sentences not appearing at the ends of transmissions had more sentence-final punctuation: 54% of text messages and 78% of IMs, e.g.,

sentence 1: I mean I just want to see you... [punctuation used]

sentence 2: I'm just stressed and overwhelmed [no punctuation]

Logically, transmission-medial punctuation is more critical than transmission-final marks in helping recipients interpret messages. In most cases, the act of sending a message coincides with sentence-final punctuation.<sup>8</sup>

To compare question marks and periods (or equivalent marks), we divided each corpus into two categories: semantic questions and "other".<sup>9</sup> More question marks were used to end semantic questions than periods (or equivalents) to end other sentence types. In texting, 73% of

semantic questions were ended with a question mark, while only 30% of “other” bore sentence-final punctuation ( $Chi^2(1) = 38.56, p < 0.001$ ). In IM, all (100%) of questions ended in question marks, while only 41% of the remaining sentences were punctuated ( $Chi^2(1) = 29.50, p < 0.001$ ). More frequent use of “required” question marks may pragmatically highlight the request for a response from the recipient.

Research Questions 1 and 2 concerned linguistic characteristics of text messaging in the USA and how they compared with American IM. Tables 1 and 2 summarize the points of similarity and difference presented in this study.

### **INSERT TABLES 1 AND 2 HERE**

The paucity of emoticons and acronyms in both our texting and IM corpora is consonant with studies in Canada and the UK. While there are no previous reports on sentential punctuation in texting or IM, our data indicate that usage patterns are hardly scattershot. Students often omitted transmission-final marks (especially periods), but their overall punctuation choices tended to be communicatively pragmatic. The fact punctuation was consistently more prevalent in IM than in texting probably reflects greater ease of input in IM.

American college-student texting and IM differed in several significant ways. Text messages were consistently longer and contained more sentences, probably resulting from both differential costing structures and the tendency of IM sequences (but not texts) to be sent seriatim. Text messages contained significantly more abbreviations than IMs, but even the number in texting was small. Reanalyzing our data to more closely approximate Thurlow and Brown’s scoring rubric, the American texting corpus contained less than 5% abbreviated words, compared with Thurlow and Brown’s nearly 19% for a sample of mostly British female college students in Wales.

Our texting and IM data also diverged with respect to contractions and apostrophes: more contractions appeared in texting, but texting used only one-third the apostrophes found in IM. Greater use of contractions in texting could reflect the higher tendency to use abbreviated forms (compared with IM), which in turn is consonant with an awkward input device. Paucity of apostrophes in texting undoubtedly results from input complexity.

Research Question 3 asked whether prior experience with word processing and IM is reflected in American texting. The answer is unclear, especially in light of differences between input devices. Many Americans have little practice texting on phone keypads. Thurlow and Brown's participants sent messages twice as long as American texts (14 words vs. 7.7; 65 characters vs. 35), which may reflect their students' more extensive texting experience. Alternatively, longer British texts may result from limited access to IM (where seriatim transmissions commonly exceed 14 words) or cultural differences.<sup>10</sup>

Explanations for uncontracted versus contracted forms in American texting are not self-evident. The fact uncontracted forms were used at all (13 out of 85 potential contractions) might reflect keyboard writing habits, as in IM, or intentional adoption of more formal written style for particular text messages. Alternatively, uncontracted forms might be used to avoid the complexity of creating apostrophes. The resulting formal expression would, essentially, be a typing shortcut. Apostrophe use is also not simply explained. Insertion of 23 apostrophes (out of 72 contractions) might reflect computer keyboard experience. However, Thurlow and Brown also reported use of apostrophes in their data,<sup>11</sup> so the issue may be general knowledge of punctuation rules, regardless of input medium.

In the future, focus groups with college students would enhance our understanding of how students craft text messages (e.g., intentionally using punctuation to make a good

impression, consciously using or avoiding abbreviations). Within the USA, longitudinal studies of both texting and IM would indicate how texting evolves with practice, and how its relationship to IM shifts. Outside the USA, it would be instructive to track how texting is affected by increased use of IM and word-processing outside North America.

### NOTES

1. International Telecommunications Union. Retrieved on February 2, 2007, from <http://www.itu.int/ITU-D/icteye/Indicators/Indicators.aspx>.
2. Internet World Stats. Retrieved on February 2, 2007, from <http://www.internetworldstats.com/stats.htm>.
3. *Le Monde* (January 9, 2005) reported an IM craze in France. About one-third of young people aged 12 to 25 were said to be using it, largely via MSN Messenger.
4. Other linguistic features we explored, such as ellipses, are omitted here for brevity.
5. Approximately one-third of the participants reported using predictive texting software.
6. Calculating sentential units per transmission entails deciding which elements to count as full sentences (e.g., *ok, haha, whatever, seriously*) and which to count as components of a preceding or following sentence (e.g., *hey Jane, besides, hon. [=honey]*). Capitalization, commas, and periods are not always reliable guides. Our criterion was whether the element commonly appeared as a stand-alone word (or phrase) in the speech of young adults.
7. Squires (2005) reported a similar finding in her study of college-student IM. In popular IM programs, there is no spell-check function that automatically inserts apostrophes.
8. However, in IM, not all transmissions conclude with the end of a sentence – see Baron, in press.

9. In texting, 18.6% of all sentential units were questions, compared with 9.9% of IMs ( $Chi^2(1) = 8.44, p = 0.004$ ). Nearly 90% of texting questions involved coordination of activity, compared with 10% of IM questions.

10. Norwegian texts averaged 6.9 words and 29-46 characters, depending upon whether they used predictive texting (Ling, 2005a).

11. Thurlow and Brown do not distinguish between apostrophes used in contractions versus possessives.

## REFERENCES

- Baron, N.S. (2003). Language of the Internet. In A. Farghali (Ed.), *The Stanford handbook for language engineers* (pp. 59-127). Stanford, California: CSLI.
- Baron, N.S. (2004). See you online: Gender issues in college student use of instant messaging. *Journal of Language and Social Psychology, 23*, 397-423.
- Baron, N.S. (in press). Discourse structures in instant messaging: The case of utterance breaks. To appear in S. Herring (Ed.), *Computer-mediated conversation*. Cresskill, NJ: Hampton Press.
- Baron, N.S., & Ling, R. (in preparation). Emerging patterns of American mobile phone use: Electronically-mediated communication in transition.
- Döring, N. (2002). "Kurzm. wird gesendet" – Abkürzungen und Akronyme in der SMS-Kommunikation. *Muttersprache. Vierteljahresschrift für deutsche Sprache, 2*.
- Hård af Segerstad, Y. (2002). *Use and adaptation of written language to the conditions of computer-mediated communication*. Doctoral Dissertation, Department of Linguistics,

- Göteborg University, Göteborg, Sweden. Retrieved February 16, 2007, from [http://www.ling.gu.se/%7eylvah/dokument/ylva\\_diss.pdf](http://www.ling.gu.se/%7eylvah/dokument/ylva_diss.pdf).
- Lenhart, A., Madden, M., & Hitlin, P. (2005). Teens and technology. Pew Internet & American Life Project, July 27. Retrieved February 16, 2007, from [http://www.pewinternet.org/pdfs/PIP\\_Teens\\_Tech\\_July2005web.pdf](http://www.pewinternet.org/pdfs/PIP_Teens_Tech_July2005web.pdf).
- Ling, R. (2005a, October). *The length of text messages and the use of predictive texting: Who uses it and how much do they have to say?* Paper presented at the Association of Internet Researchers, Chicago, IL.
- Ling, R. (2005b). The socio-linguistics of SMS: An analysis of SMS use by a random sample of Norwegians. In R. Ling & P. Pedersen (Eds.), *Mobile communications: Re-negotiation of the social sphere* (pp. 335-349). London: Springer.
- Ling, R., & Haddon, L. (in press). Mobile emancipation: Children, youth, and the mobile telephone. To appear in K. Dortner & S. Livingstone (Eds.), *International handbook of children, media, and culture*. London: Sage.
- Squires, L. (2005, October). *Whats the use of apostrophes? Gender differences and linguistic variation in instant messaging*. Paper presented at the Association of Internet Researchers, Chicago, IL.
- Tagliamonte, S., & Denis, D. (in preparation). LOL for real! Instant messaging, teen language and linguistic change.
- Thurlow, C., & Brown, A. (2003). Generation Txt? The sociolinguistics of young people's text-messaging. *Discourse Analysis Online*. Retrieved February 16, 2007, from <http://extra.shu.ac.uk/daol/articles/v1/n1/a3/thurlow2002003-paper.html>.

Traugott, M., Joos, S.-H., Ling, R., & Qian, Y. (2006). *On the move: The role of cellular communication in American life*. Pohn Report on Mobile Communication. Ann Arbor, MI: University of Michigan.

## BIOGRAPHIES

Rich Ling (Ph.D., University of Colorado) is a senior researcher at Telenor R&D and has an adjunct position at the University of Michigan. His interests include the social consequences of mobile telephony. He is author of *The mobile connection*, and along with Per Pedersen has co-edited *Mobile communications: Re-negotiation of the social sphere*. Ling's forthcoming book is *Mediated ritual communication*.

Naomi S. Baron (Ph.D., Stanford University) is a professor of linguistics at American University in Washington, DC. Her research interests include computer-mediated communication, the influence of technology on spoken and written language, and the history and structure of English. Author of *Alphabet to email: How written English evolved and where it's heading*, her forthcoming book is *Always on: Language in an online and mobile world*.

Table 1  
*Similarities between American Text Messaging and IM*

<u>Feature</u>	<u>Texting</u>	<u>IM</u>
<b>Emoticons and lexical shortenings</b>		
emoticons	.001% of words	.004% of words
acronyms	.005% of words	.003% of words
<b>Sentence punctuation</b>		
overall sentence punctuation	39% of sentences	45% of sentences
transmission-final punctuation	29% of sentences	35% of sentences
transmission-internal punctuation	54% of sentences	78% of sentences
use of required question mark	73% of questions	100% of questions
use of required period	30% of other sentences	41% of other sentences

Table 2  
*Differences between American Text Messaging and IM*

Feature	Texting	IM
Length		
transmissions (in words)	7.7 words	6.0 words
transmissions (in characters)	35 characters	29 characters
one-word transmissions	3.7% of messages	18.8% of messages
multi-sentence transmissions	60% of messages	34% of messages
sentences per transmission	1.76 per transmission	1.27 per transmission
Emoticons and lexical shortenings		
abbreviations	3.2% of words	0% of words
contractions	84.7% of potential	68.1% of potential
apostrophes	31.9% of contractions	93.9% of contractions