

Association for Information Systems AIS Electronic Library (AISeL)

AMCIS 2004 Proceedings

Americas Conference on Information Systems
(AMCIS)

December 2004

Recovering Punctuations in Instant Messages - Towards the prosody norm in IM

Lina Zhou
University of Maryland

Ziming Zhuang
University of Maryland, Baltimore County

Follow this and additional works at: <http://aisel.aisnet.org/amcis2004>

Recommended Citation

Zhou, Lina and Zhuang, Ziming, "Recovering Punctuations in Instant Messages - Towards the prosody norm in IM" (2004). *AMCIS 2004 Proceedings*. 408.
<http://aisel.aisnet.org/amcis2004/408>

This material is brought to you by the Americas Conference on Information Systems (AMCIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in AMCIS 2004 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

Recovering Punctuations in Instant Messages

— Towards the prosody norm in IM

Lina Zhou

Department of Information Systems
University of Maryland, Baltimore County
zhoul@umbc.edu

Ziming Zhuang

Department of Information Systems
University of Maryland, Baltimore County
rick.zhuang@umbc.edu

ABSTRACT

Instant Messaging (IM) is being increasingly adopted in the workplace and personal life. Prosody can enhance users' interaction with IM. Punctuation is one type of indicators of prosody in text. However, instant messages are characterized by absent punctuations, among others. Moreover, the empirical study on punctuations in IM is sparse. Therefore, we aim to recover punctuations in instant messages. In this study, we develop a heuristic approach to automatically recovering punctuations in IM. The empirical results of the approach are very encouraging. The findings of this study have a number of implications. They shed light on the patterns of missing punctuations in instant messages and the importance of proper punctuation usage to achieving effective communication. This study is the first effort toward creating the prosody norm in IM interaction.

Keywords

Instant Messaging, prosody, punctuation, recovery, natural language processing.

INTRODUCTION

Instant Messaging (IM) is a system that supports computer-based synchronous or near-synchronous point-to-point communication between different users over the network. The fast pace of message exchanges and the flexible support of multimedia communications in different settings has made IM one of the most widely used communication tools in the community. Gartner predicts that the total number of IM users will grow to 180 million worldwide by 2004. IM has evolved from an informal interpersonal communication medium for keeping in touch with friends and family, to an online platform for facilitating effective coordination, scheduling tasks and event in enterprises, and managing customer relationships. The percentage of IM accounts owned by business users is expected to increase to 50 percent by 2004 (Fordahl, 2002). Such a shift in context of IM use poses challenges to traditional IM practice. One of them is related to the language style.

Instant messages in general are featured with informal and spoken language style. It is exemplified in absent punctuations, lacking upper cases, special abbreviations, tolerance of grammatical errors, unconventional use of punctuations in expressing emotions, and so on. All of the above features are in line with the fast pace of synchronous communication. However, they are achieved at the cost of sacrificing prosody in IM. To our best knowledge, the importance of prosody in IM has rarely been investigated.

Prosody is commonly associated with acoustic characteristics of speech utterance, which include the pitch or frequency, the length or duration, and the loudness or intensity. The varying quantities in the above forms help determine the function to which listeners orient themselves in interpreting the utterance (Peppé and Maidment, 2000). In view of the similarity between IM and spoken communication, we believe that prosody can also play an important role in IM to help receivers interpret incoming messages, as shown in speech conversation (Peppé and Maidment, 2000). Prosody can enhance the interaction with IM in some important ways. Prosody in IM interactions may indicate syntax chunks, turn-taking, types of messages such as questions and statements, and sender's attitudes and feelings. Among a variety of indicators of prosody in IM, we focused on punctuations in this study.

This study aims to achieve two objectives. One is to examine the role of punctuations and the impact of recovering punctuations in IM. The other objective is to develop an approach to effectively recovering missing punctuations in instant messages.

We believe that the findings of this study can benefit researchers, public users, service providers and tool developers of IM. First, it integrates linguistics, human language technologies, and computer-mediated communication research in examining punctuations in IM. Second, it provides guidance to broader IM users in properly using punctuations in IM. Third, the findings provide suggestions for IM providers to improve system interface and functionalities in order to attract and retain users.

This rest of paper is organized as follows. We first review the role of prosody in IM and the implications of automatic recovery of punctuation in IM. Then, we present a heuristic approach to automatically recovering punctuations in IM. Next, we design a study to empirically evaluate the proposed approach. Finally, we discuss the research findings and suggest research questions for further investigation.

BACKGROUND

The Role of Prosody in Instant Messaging (IM)

Previous studies have examined the characteristics of language styles in IM conversations. Informality is the predominant feature of IM messages with highly colloquial grammar, exaggerating or ignoring punctuations, non-standard abbreviations and relaxed spelling (Nardi et al. 2000). The family or friend relationship between communication partners may have led to the relaxed and informal conversation style. As IM enters workplace in dealing with work-related tasks and events, IM interaction becomes less informal so as to prevent any possible misunderstanding. Prosody may enhance the effectiveness of IM communication by marking syntax chunks, types of messages such as questions and statements, and so on. The following major observations of IM communication (Connell et al. 2001; Isaacs et al. 2002; Radical Group, 2001) shed light on the importance of prosody in IM.

- Instant messages are usually short. Due to the synchronous or near-synchronous nature of IM, most of the IM interactions are dealing with a single purpose such as coordinating schedules, asking questions and offering answers. Short messages themselves do not offer rich context for interpretation. Harnessing prosody in instant messages may complement the limited context in short messages.
- Frequent media switching is likely to occur during the conversation. Users may switch to other media when: 1) the conversation becomes too complex to continue in IM; and 2) IM conversation is initiated to arrange an interaction with other media (Nardi et al. 2000). Both of the scenarios imply that IM is a lean medium, whose capacity inhibits users from fully expressing or understanding information. Thus, it is desirable to maintain prosody in instant messages, which may enhance the expressivity of IM.
- Users are often multitasking while they are chatting. Users are likely to engage in other tasks, such as checking email, surfing the web, or even carry on multiple simultaneous conversations while they are chatting with one another. When switching between multiple applications, a user may lose the track of who is who and misuse communication styles in message exchanges. As a result, misunderstanding may cause serious consequences in business or personal relationships. Even if IM provides preceding context for disambiguation in ongoing interaction, frequent reference to local context may compromise the rapid pace of IM interaction. The above problem may be mitigated if prosodic information is available.

In sum, prosody plays an important role in enhancing the effectiveness of IM interaction. We selected an important indicator of prosody in IM - punctuations – for examination in this study.

Implications of Recovering Punctuations in IM

We focus on punctuations in the current study. The significance of punctuations to IM is not only reflected in the role of prosody in IM interaction, but also observed from many other perspectives, including linguistics, natural language processing, text-to-speech generation, and so on. The disparity between the implication of punctuations to IM and the IM practice of frequently ignoring punctuations highlights the importance of recovering punctuations in IM.

Linguistics Perspective

Instant messages are written but they are very close to spoken language (e.g., synchronicity). In instant messages, punctuations not only reveal “how writers view the balance between spoken and written language (Baron 2000)”, but also convey the prosody – and they argue that it’s the only way to do so – of the messages, which in spoken language can be expressed by different tones and various length of pauses during the conversation. Nevertheless, the absence of punctuations, when they should have been used, is common in IM conversations.

Crystal (2001) observed that “Punctuations tends to be minimalist in most situations, and completely absent in some emails and chat exchanges.” Crystal (2001) also recognized the importance of punctuation usage: “it is the chief means a language has for bringing writing into direct contact with (the prosody and paralinguistic of) speech, as well as conveying a great deal of information about grammatical construction.” For example (Nunberg, 1990):

“He reported the decision: we were forbidden to speak with the chairman directly.”

“He reported the decision; we were forbidden to speak with the chairman directly.”

In the above examples, different punctuations convey distinctively different meanings. In the first case, the clause after the colon represents the content of *decision*, while the same clause in the second sentence is considered relatively independent of *decision*. Therefore, ignoring punctuations may result in mis-interpretations of a message, which calls for the recovery of punctuations.

Natural Language Processing (NLP) Perspective

Although originally most NLP systems, mostly designed for parsing and discourse analysis, did not take punctuations other than blank spaces and periods into account, many recent studies emphasize the usage of punctuations in syntactic and semantic analysis. In addition to serving as delimiters between grammatical and structural units, punctuations are also used for separating and distinguishing morphology, and resolving ambiguities (Say and Akman 1998). Jones (1995) argued that the use of punctuation within NL systems could be beneficial because disambiguating marks, usually commas, occur where an unintentional ambiguity could result if the marks were not there. The statement was empirically supported by comparing the parsing results of the same corpus using two similar grammars with one involving punctuations and the other not. It was found that taking punctuations into account helped the parsing processing to various degrees, with regard to different complexity of sentences (Jones 1995). Some studies focused on certain types of punctuations. For example, Bayraktar et al. (1998) investigated the usage of commas by building syntactic rules to facilitate computer-aided classification of comma usages in a given parsed corpus. Say and Akman (1997) provided a detailed review on current approaches to computer-aided linguistics study, and proved punctuation’s significance from an information-based perspective.

Text-to-Speech Perspective

Punctuations are critical to expressing the prosody and tone in converting text messages into synthesized voice. Beeferman et al. (1998) proposed an approach to the automatic insertion of punctuation in a speech dictation system. By means of accurate and efficient punctuation restoration, the system achieved better performance in qualitative measures (e.g., user satisfaction). We envision that text-to-speech technologies can be integrated into IM to facilitate computer in generating speech from text messages. Punctuations in instant messages would be extremely helpful to producing natural speech by restoring the prosody and tone of the messages. For example, a comma in text can be signaled by a pause in speech, which helped to explain “how in written language we attempt to communicate some of the suprasegmental aspects of speech through punctuation (Edwards 1994).” Edwards (1994) argued that “the quality of a voice is judged on both its segmental fidelity and its suprasegmental features”, which further suggests that the text input to the voice synthesizer should include punctuations.

It is suggested from the above analysis that punctuations could be essential to accurately expressing and interpreting instant messages. However, missing punctuations are still prevalent in IM. In order to improve the effectiveness of IM interaction, we aim to recover punctuations in IM. Moreover, for the sake of efficiency and scalability, we propose an approach to automatically recovering punctuations.

METHODOLOGY DESIGN

We proposed a heuristic approach to automatically recovering punctuations in instant messages. The approach was based on both structural and linguistic knowledge. It was evaluated with actual instant messages collected from an experiment study. Messages were manually checked for missing punctuations, which were inserted as appropriate. The manual recovery was performed by two subjects separately and their results were cross-validated. The performance of the heuristic approach was evaluated by comparing the result of the automatic recovery against that of manual validation.

Heuristic Rules

The heuristic approach mainly consists of heuristic rules. There are two types of heuristic rules: message structure rules and punctuation recovery rules. They are described in formal language as follows:

Message Structure Rules

Let *msg* denote a single message; *st* denote a grammatically complete sentence; *cls* denote a clause that contains a subject and a predicate and is a part of a sentence; *np* denote a non-verbal phrase that does not contain any verb; $(\cdot)^*$ represents the Kleene closure of \cdot ; and *f* indicate sentence fragments. WH (WH-words and how), BE (Be-verbs), MD (modal verbs), DO (auxiliary verb do), V (regular verbs), R (pronouns) and C (conjunctions) indicate a word class (i.e., part-of-speech) respectively. For example, the first rule shows that a message is the concatenation of one or more strings of an element and zero or more punctuations. The second rule shows that an element consists of a sentence, a clause, or a non-verbal phrase.

$msg \rightarrow elem (punc)^* (elem (punc)^*)^*$
 $elem \rightarrow st | cls | np$
 $st \rightarrow (cls | (cls)^*)$
 $st \rightarrow (inte\ be) (cls | (cls)^*)$
 $punc \rightarrow , | . | ? | ! | f$
 $inte \rightarrow WH | auxv$
 $auxv \rightarrow BE | MD | DO$

Punctuation Recovery Rules

Once a message is parsed with structure rules, missing punctuations in the message can be restored with punctuation recovery rules. Sample rules are described as follows:

Suppose $msg = w_1w_2...w_l..w_m..w_n$, where w_l is the first word following the last punctuation in message *msg*; ($w_l = w_l$ when there is no punctuation in *msg*); $pos(w)$ is the part-of-speech of w ; w_{n+1} is the first word of the message following *msg*; and *RP* indicates recovered punctuations. In addition, we use w_0 to represent the last word in the preceding message.

- Rule 1: $pos(w_l) \in inte \ \& \ pos(w_{l+1}) \in auxv \Rightarrow RP = '?'$
- Rule 2: $pos(w_l) \in auxv \ \& \ (pos(w_{l+1}) \in np \ | \ pos(w_{l+1}) = 'R') \Rightarrow RP = '?'$
- Rule 3: $pos(w_{n+1}) = 'C' \Rightarrow RP = ','$
- Rule 4: $\forall i, pos(w_i) \neq 'V' \ \& \ w_0 \neq '?' \Rightarrow RP = 'f'$
- Rule 5: $\Rightarrow RP = '.'$

Rule 5 is a default one, which has no antecedent part.

Algorithm

We implemented the heuristic rules in C#, which ran on Microsoft .NET framework. The rules are activated in the descending order of their specificity. Matched rules with the strictest constraints are executed first. The algorithm that applies the sample rules to recover punctuations is described in Figure 1.

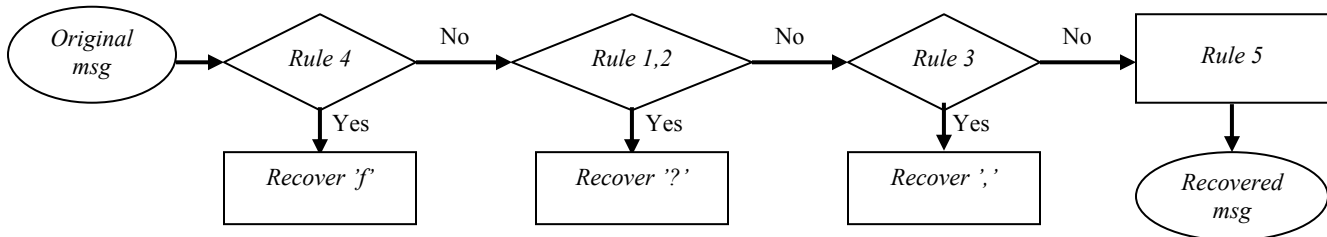


Figure 1. Algorithm for Recovering Punctuations

Evaluation Metrics

The heuristic approach to recovering punctuations was evaluated with two measures: a) precision rate, which is defined as the percentage of recovered punctuations that are correct; and b) recall rate, which is defined as the percentage of missing punctuations that are correctly recovered.

DATA COLLECTION AND EVALUATION

Experiment Design for Data Collection

Participants were randomly formed into 2 or 3-person groups to conduct a task with IM. Each group was assigned a hypothetical decision making task. Participants were asked to discuss in groups through exchanging instant messages. The discussion session was stopped when group members finish the task or they ran out of time of 15 minutes. They were not allowed to use any other applications during the experiment.

Sixty-seven subjects participated in this study. Participants were undergraduate students recruited from two information systems courses in a mid-sized university on the east coast. They were all junior and seniors. All of them have experience with IM. In order to avoid the effect of personal familiarity on the style of instant messages, participants in the same groups did not see one another before and during the experiment. In addition, participants used pseudo-names, which were assigned at the site, in the IM discussion.

Data Preparation

Text messages from all the sessions were extracted from the archived log files, which preserved the original punctuations as they were originally typed in. 1758 instant messages were collected. Among them, 1270 messages (72.241%) had no ending punctuations. There were another 11 (0.6%) messages that had missing punctuations in the middle. In view of the small number of messages in the latter case, we temporarily ignored them in this study.

Two graduate students, who were experienced IM users, were recruited to manually recover missing punctuations independently. The inter-rater reliability analysis of punctuation recovery showed that Cronbach's alpha value was greater than 0.7. They disagreed on 47 out of 1270 messages (3.701%) in terms of punctuation. To ensure objectivity and validity of the evaluation, we only used those messages whose punctuations were recovered consistently by the two subjects.

RESULTS

The manual and automatic recovery results are displayed in Figure 2. The precisions and recalls of the heuristic approach for different types of punctuations were tabulated in Table 1. It is shown that all types of punctuations except commas were recovered with satisfactory results. The breakdown of errors in automatic recovery is also listed in Table 1.

Types	Number of Correct Recovery	Recall	Precision	Type / Number of Errors in Recovery
Period	887	0.92880	0.95582	Period → NP: 32 Period → Comma: 32 Period → Question: 4
Comma	64	0.65979	0.66667	Comma → NP: 0 Comma → Period: 33 Comma → Question: 0
Question Mark	67	0.87013	0.94366	Question → NP: 3 Question → Comma: 0 Question → Period: 7
Non-verbal Phrase (NP)	93	0.98936	0.72656	NP → Period: 1 NP → Comma: 0 NP → Question: 0
Overall	1111 out of 1223 (90.842%) messages were correctly recovered.			

Table 1. Detailed Summary of the Results of Punctuation Recovery

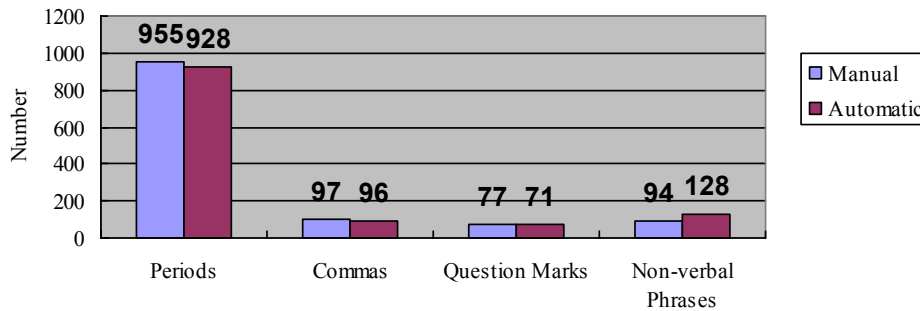


Figure 2. Comparison of Manual and Automatic Results

DISCUSSION

Error analysis

In this section, we analyze the types of errors generated by the automatic approach. It provides insight into language ambiguity introduced by missing punctuations in instant messages as well as the strengths and weaknesses of the proposed heuristic approach to punctuation recovery in IM.

Period → NP: Context and Grammar

32 periods (3.35% of missing period cases) were mistakenly recognized as non-verbal phrases. Such errors were mainly caused by two factors. The first is the relaxed spelling and extremely colloquial style of instant messages, such as “oh i c”, “k, cya”, and so on. Letters such as ‘c’ and ‘k’ represent words ‘see’ and ‘okay’ respectively. They are not compiled into the generic dictionaries. The second is the lack of context, such as “23 miles a day”, “second here”, and so on, which were actually short answers to previous questions. They were not identified due to cascading errors from omitted questions or dynamic turn-takings during IM interactions. Different participants have the flexibility of sending messages in parallel in IM. As a result, an answer may be several messages away from the original question.

Period → Comma: Rule and Grammar

Misrecovering periods with commas accounted for 32 errors (3.35% of missing period cases). The primary reason for this type of errors is that the heuristic rules did not consider the factor of topic change. A topic shift is more likely to indicate initiating a new sentence than post-positioning a clause. This issue may also be compounded with relaxed grammar. For example,

“we could use the lens of the glass with the sun rays and creat a fire on the tire like dat,”
“so what is the closest town or minning camp?”

In the above case, even though the second message is initiated with a conjunction ‘so’; it is considered as an independent sentence. In a strict grammar sense, ‘so’ should not be included in the first place.

Period → Question: Context and Grammar

There were 4 errors in this category. The analysis of error cases revealed that it is difficult to differentiate a standalone non-restrictive relative clause from a question statement. For example,

“if we stay put we canhave shelter.”
“not if we get rescued first.”
“which is why we should set up camp and wait it out?”

In the above case, our approach may fail to recognize the relationships between the last message and the first two messages. Consequently, the last message was recover with a question mark. Again, understanding a large context is extremely crucial to this case. It is interestingly noted that IM users tend to express a complete statement with several separate and consecutive messages. They involve complex ellipses and reference resolution issues, which still challenges NLP communities.

Comma → Period: Context

Another 33 errors (34.02% of missing comma cases) were results of misrecognizing commas as periods. Most of them involved subordinate-clause messages, which should be attached to main clauses in either the preceding or the following messages. For examples:

“you could use the jack to lift a rock.”
 “if it is too heavy,”

The if-clause should be attached to the preceding message to form a complete sentence in the above example. Such ambiguity in attaching subordinate clauses could be resolved by semantic and discourse analyses. Without deeper understanding of the local context, the current approach failed to identify the relationship between ‘it’ and ‘rock’ and between ‘heavy’ and ‘rock’.

Question → NP: Grammar

There were 3 errors in this category. They were mostly caused by relaxed spelling and grammar. For example,

“whats left.”
 “y not.”

The above two messages would have been recovered with question marks if they were fully spelled out as “what’s left” and “why not” respectively. This kind of unconventional short notation and relaxed spelling are beyond the scope of traditional language analysis tools.

Question → Period: Grammar & Context

Among the messages with missing question marks, 7 were mistakenly recovered with periods. Such errors were mainly caused by ellipses in interrogative sentences. In the following messages, for example,

“why you need iodine tabs for.”
 “sweet strap or bag next.”

In the first message, ‘do’ in “why do” was omitted. Thus, it was considered as a clause rather than an interrogative sentence. The initiation of a question, “Do we select”, was entirely absent in the second message. As a result, neither of the messages received correct recovery. The local context may also be helpful in correctly recovering the above punctuations.

There were other types of errors in addition to the ones discussed above. Due to space constraint, we do not enumerate them here. After detailed analysis of misrecognized punctuations, we summarize the causes of recovery errors and their potential solutions into three categories: 1) lack of **context**, which could be solved by taking larger context into account; 2) relaxed spelling and **grammar**, which may be addressed by training statistical language models on corpus of instant messages; and 3) exception to the **rule**. The heuristic rules for recovering punctuations in instant messages are still evolving. Existing rules can be modified or extended to handle exceptional cases. Table 2 summarizes the error types and their potential causes and solutions.

Error Types	Context	Grammar	Rule
Period → NP	X	X	
Period → Comma		X	X
Period → Question	X	X	
Comma → Period	X		
Question → NP		X	
Question → Period		X	X

Table 2. Types of Errors in Punctuation Recovery and their Causes and Potential Solutions

Patterns of Missing Punctuations

The observations and statistics on the experiment data revealed two major patterns of missing punctuations. One is that, punctuations at the end of messages were most likely (99.141%) to be ignored. This does mean that only punctuations at the end of sentences were missing, for a long sentence tends to be broken down into multiple short segments and sent in a sequence of messages. We believe that breaks between consecutive messages correspond to pauses in speech, which attach rhythmic prosody to the messages being sent. Second, punctuations are more likely to be omitted in descriptive sentences, for 86.02% of the missing cases were involved with periods or commas. Non-descriptive sentences, such as interrogative sentences, express special tones or emotion, which are important indicators of prosody in instant messages. Both of the above patterns show that prosody was preserved in IM in some sense. However, a significant number of punctuations were still lost in instant messages, which motivated the current research.

Research and Practical Implications

We can make three important observations from the results of automatic punctuation recovery and error analysis. The findings have both research and practical implications to IM communication.

- Among the three major types of causes of errors, grammar (relaxed spellings and grammars) is the predominant factor. The collected messages provide empirical evidence showing that users tend to care less about grammar conventions in IM communication. For example, capital cases are rarely available for identifying the initiation of a sentence or special phrases. It makes traditional rule-based NLP tools ineffective in processing ungrammatical instant messages. We believe that a customized tool should be developed for parsing instant messages. Such a tool should incorporate special acronyms, frequent ellipses, and partial sentences in analyzing a message. Compared with the rule-based approach, statistical approach (Charniak 1997) may be more robust in dealing with ungrammatical structures or unconventional spellings.
- Context is critical to recovering punctuations in instant messages. Semantic and discourse analyses using contextual information remain as challenging issues in NLP. Moreover, it is both cognitively demanding and time consuming for a user to refer to old messages in interpreting an incoming message with absent punctuations. Therefore, the automatic recovery of punctuations is a promising solution to enhancing the efficiency of interpersonal communication with IM. It appears that punctuations are not equally important. For example, if a message consists of a non-verbal phrase (e.g., a short answer in response to a preceding question), it should be clearly marked with a period. If a message consists only of a subordinate clause, it is better to use the punctuations to indicate whether the message should be attached to the previous message (using period), or merged with the following message (using comma).
- The unconventional patterns of composing one sentence with multiple messages and connecting multiple sentences with conjunctions to indicate a logical inference path pose new challenges to punctuation recovery. Ignoring capital cases in instant messages makes the problem even more challenging. Similar problems are encountered in speech synthesis (Beeferman, et al 1998) and text segmentation communities (Ponte and Croft 1997). Methods developed by relevant disciplines may help address the problems in IM.

The findings can potentially be used to developing customized NLP tools and building knowledge resources for understanding instant messages, designing more user-friendly interface for IM tools, and supporting text-to-speech generation in IM communication. Moreover, this study may lead to creating prosody norms for IM interaction and increasing users' awareness of the importance of preserving punctuations in IM communication.

Limitations

Cautions should be given when extending the patterns discovered in this study to other contexts. First, participants in the experiment were peers, but they did not know whom they were interacting with during IM discussion. If they were grouped based on the friend relationship, they may chat in a more informal way and omit more punctuation. Their acquaintance history may help them interpret messages by reading prosody between lines. In contrast with peers, subordinate-superior relationships in hierarchical organizations or business-customer relationships usually require more formal interaction style with prosody clearly expressed in order to avoid any misunderstanding. In addition, our approach was mainly based on syntactic information, which may be improved by incorporating semantic and contextual information.

CONCLUSIONS AND FUTURE DIRECTION

We discussed the importance of automatic recovery of punctuations in instant messages from the prosody perspective, and developed a heuristic approach to automatically recovering punctuations. The empirical results on instant messages showed

that the proposed approach was encouraging, which correctly recovered over 90 percent of omitted punctuations. The analysis of the instant messages revealed the patterns of missing punctuations and shed lights on the significance and challenges of recovering punctuations for IM communication.

We believe that this study filled in a gap in the current IM research by addressing the issue of punctuations in instant messages. As IM becomes increasingly prevalent nowadays at both individual and organizational levels, the impact of research along this line will be significant. Directions of future studies lie in many possibilities. We plan to investigate the importance of prosody in instant messages by empirically analyzing the effect of recovered punctuations on users' satisfaction, effectiveness, and efficiency in IM communication. We will continue to examine other indicators of prosody in IM communication, such as pauses and response length.

REFERENCES

1. Baron, N. S. *Alphabet to Email*. (2000) Routledge, pp. 167.
2. Bayraktar, M., Say, B., and Akman, V. (1998) An Analysis of English Punctuation: The Special Case of Comma.
3. Beeferman, D., Berger, A., and Lafferty, J. (1998) Cyberpunc: A Lightweight Punctuation Annotation System for Speech.
4. Charniak, E. (1997) Statistical Techniques for Natural Language Parsing, *AI Magazine* 18(4): 33-44.
5. Connell, J. B., Mendelsohn, G. A., Robins, W. A., Canny, J. (2001) Effects of Communication Medium on Interpersonal Perceptions: Don't Hang up on the Telephone Yet! Presented at *GROUP 2001*.
6. Crystal, David. (2001) *Language and the Internet*, Cambridge: Cambridge University Press, pp. 89 – 90.
7. Edwards, A. (1994) *Speech Synthesis, Information Technology and Disabilities*, Vol 1. No. 2.
8. Fordahl, M. (2002) Instant message monitoring soaring, *The Sydney Morning Herald*.
9. Isaacs, E., Walendowski, A., Whittaker, S., Schiano, D. J., and Kamm, C. (2002) The Character, Functions, and Styles of Instant Messaging in the Workplace, *Proceedings of the 2002 ACM Conference on Computer Supported Cooperative Work*, 2002, New Orleans, Louisiana, USA, pp. 11 – 20.
10. Jones, B. (1994) Exploring The Role of Punctuation in Parsing Natural Text.
11. Nardi, B., Whittaker, S., Bradner, E. (2000) Interaction and Outeraction: Instant Messaging in Action, *Proceedings of the 2000 ACM Conference on Computer Supported Cooperative Work*, Philadelphia, Pennsylvania, United States, pp. 79 – 88.
12. Nunberg, G. (1990) The Linguistics of Punctuation, *CSLI Lecture Notes*, No. 18, University of Chicago Press.
13. J.M. Ponte and W.B. Croft. (1997) Text Segmentation by Topic, *Proceedings of the First European Conference on Research and Advanced Technology for Digital Libraries*, pp. 120-129.
14. Radical Group. (2001) Instant Messaging and SMS: Market Trends 2001-2004, *Radical Market Report*, 2001.
15. Say, B. and Akman, V. (2001) Current Approaches to Punctuation in Computational Linguistics.
16. Smith, David. (2001) Tackle Instant Messaging's Security Risks – Now, *Internet Strategies*, Gartner Inc. <http://www3.gartner.com/resources/101600/101608/101608.pdf>. 2001.
17. Peppé, S. and Maidment, J. (2000) Prosody on the Web, <http://www.eptotd.btinternet.co.uk/pow/powin.htm>.