Same Translation but Different Experience: The Effects of Highlighting on Machine-Translated Conversations

Ge Gao[1], Hao-Chuan Wang[3], Dan Cosley[2], Susan R. Fussell[1,2]

[1]Department of Communication
[2]Department of Information Science
Cornell University
Ithaca NY 14850 USA
[gg365, drc44, sfussell] @cornell.edu

[3]Department of Computer Science &
Institute of Information Systems and Applications
National Tsing Hua University
101, Sec.2, Kuang-Fu Rd. Hsinchu 300, Taiwan
haochuan@cs.nthu.edu.tw

ABSTRACT

Machine translation (MT) has the potential to allow members of multilingual organizations to interact via their own native languages, but issues with the quality of MT output have made it difficult to realize this potential. We hypothesized that highlighting keywords in MT output might make it easier for people to overlook translation errors and focus on what was intended by the message. To test this hypothesis, we conducted a laboratory experiment in which native English speakers interacted with a Mandarin-speaking confederate using machine translation. Participants performed three brainstorming tasks, under each of three conditions: no highlighting, keyword highlighting, and random highlighting. Our results indicated that people consider the identical messages clearer and less distracting when keywords in the message are highlighted. Keyword highlighting also improved subjective impressions of the partner and the quality of the collaboration. These findings inform the design of future tools to support multilingual communication.

Author Keywords

Multilingual communication; brainstorming; machine translation; highlighting

ACM Classification Keywords

H.5.3 [Group and Organization Interface]: Computer-supported cooperative work

General Terms

Experimentation; Human Factors

INTRODUCTION

Globalization, with the increasingly widespread use of the Internet, has created more opportunities for people to interact with others who speak different native languages. Multilingual organizations often choose to use a common language (*lingua franca*), such as English, and provide intensive language training for all employees [9]. Using a common language reduces the need for expensive human or machine translation. However, speaking a common language can have negative consequences for non-native speakers. Non-native speakers may fear speaking up when they have less than perfect fluency in the common language of the organization [10][22]. They may also splinter into subgroups, each speaking a different native language [e.g., 23]. At the same time, native speakers may be hesitant to converse with non-native speakers because of concerns about their addressee's fluency [e.g., 3].

In recent years, machine translation (MT) technology has made it possible, in principle, for members of multilingual organizations to interact via their own native languages [e.g., 15]. MT tools should theoretically be able to eliminate many of the problems created by use of a common language, such as concerns about fluency or subgroup splintering. However, current MT services still sometimes produce erroneous translations (e.g., by translating *computer bug* into the equivalent of *computer insect*), or by forming poor sentence compositions (e.g. by translating the Chinese sentence "*得在北京时间的六点之前把文件发出去*," equivalent to "*Need to send out this document before 6 o'clock Beijing time,*" into the English translation "*Was in Beijing on a six point document send out*").

Problems in translation quality can make it difficult for group members to establish common ground [7], particularly when teams must refer to objects and entities in a workspace. As a result, studies have shown that when communication requires coming to agreement on objects of reference, using MT is less efficient than using a shared second language [28][30].

Although improvements to MT technology will continue to be made, it will likely be a long time until complete and accurate translations can be made between all languages due to the diversity and complexity of human languages. The accuracy of MT can vary with the length of a sentence, the context of use, and the specific language used [2][18].

In this paper, we take an alternative approach: Rather than trying to improve the quality of MT, we consider modifying presentation of translation results to make the output more useful. We look for simple modifications that can detour

CHI 2013, April 27–May 2, 2013, Paris, France.

around the technical limitations of MT, while improving the efficacy of communication at limited cost.

The specific modification we focus on is the addition of keyword highlighting. Such highlighting can be valuable in several ways. It can improve the organization and display of information for people to process [16][25] and support message comprehension [14]. It is possible to highlight keywords in MT-mediated communication through human processing (e.g., manual annotation), by machine processing (e.g., linguistic analysis of semantically centering words [31]), or a combination of the two (e.g., MT-assisted collaboration annotation [4]). By highlighting keywords of translated messages, it may be possible to direct people's attention to words or concepts salient in a sentence and help guide people to correct interpretations, even if the translation is less than perfect. For example, if we highlight the keywords of the translated sentence "*Was in **Beijing** on a **six** point **document** **send** **out**,*" its intended meaning might be easier to discern and comprehend.

In the remainder of this paper, we first review literature suggesting that highlighting could increase the value of MT output for real-time communication and outline our hypotheses. We then present a study in which we compare keyword highlighting to two control conditions: no highlighting and random highlighting. Native English speakers performed three collaborative brainstorming tasks, one in each condition, with a Mandarin-speaking confederate. After each task, they rated their understanding of the messages, their impressions of their partner, and the quality of the collaboration. Keyword highlighting led to higher levels of understanding and more positive perceptions of partners and the collaboration versus no highlighting or random highlighting. The results can inspire the future development of MT-based communication tools.

BACKGROUND AND RELATED WORK

Textual Highlighting and Information Processing

Highlighting has been used as a strategy to support people's processing of information. With paper books, researchers noticed that readers preferred to use highlight markers and boxes to organize information on textbooks [e.g., 16]. With the development of digital media, the strategy of using highlighting has been applied to the design of email system interfaces, webpages, search tools, and web-based parallel translation systems (e.g., [1][5][6][24]).

The main function of highlighting is to help redistribute cognitive resources when processing messages. Since meaningful highlighting can improve the way information is organized and displayed, it reduces the cognitive load of information processing [25]. In an information search task, for example, highlighting the target information can improve searching and reading performance. When Web users want to track textual changes on the same web site at different times, highlighting changes can improve users' awareness of the dynamism of the web content [24].

Based on this evidence, we hypothesized that information processing during MT-mediated communication could also be directed by highlighting. The value of the highlighting, however, should differ depending on whether or not the highlighted words are semantically important for understanding the whole message. When key words in a message are highlighted, paying more attention to them should be beneficial to information processing. However, when random words in a message are highlighted, paying attention to them should not be helpful to understanding, and the highlighting can also have negative, distracting effects to information processing.

Textual Highlighting and Understanding

Highlighting may benefit people's understanding of textual material. When reading traditional textual books, doing highlighting (such as underlining) under important words helps improve understanding of the content. But this benefit disappears when unimportant words in the book content are highlighted [e.g., 13]. Kawasaki and colleagues [14] examined the effect of highlighting with computer-based reading tasks, finding that people's understanding of digital articles could be improved when important words, phrases, and sentences in the text were pre-highlighted.

In this study, we were interested in detecting whether highlighting would influence the understanding of messages in MT-mediated multilingual communication. It was worth noting that the semantic understanding under the current scenario would be different from article reading in previous research in several aspects. First, the textual messages would show up within conversation, which makes them more informal and flexible than messages in written articles. Second, the textual messages were generated by machine translation, which means the quality of the language might not be as good as texts written in one's native language. Third, in synchronous conversation, people are expected to understand and respond to others quickly. Overall, the difficulty and cognitive load in processing a machine-translated message during multilingual communication can be much higher than processing messages in an article.

Thus, we argue that the use of highlighting will be especially helpful for MT-mediated multilingual communication. Although imperfect translation between languages can be distracting, highlighting keywords of each

Figure 1. The *extra thumb* task with a brainstorming question of "What are the benefits and difficulties if people had an extra thumb on each hand".

message should help improve the clarity of messages and reduce the distraction of translation errors.

H1: People will find MT-translated messages with keyword highlighting to be more understandable than messages with no highlighting or with random highlighting.

H2: People will find MT-translated messages with keyword highlighting to be less distracting than MT-translated messages with no highlighting or random highlighting.

Random highlighting, on the other hand, could diminish clarity and introduce distractions because it focuses people's attention on possibly irrelevant portions of a message. Alternatively, people might ignore random highlighting altogether after determining it had no value for comprehension. We thus posed the following two research questions:

RQ1: How will random highlighting of a message affect the clarity of a message relative to the same message without any highlighting?

RQ2: How will random highlighting of a message affect ratings of distraction relative to the same message without any highlighting?

Textual Highlighting and Social Experience

In addition to information processing and understanding, we were also interested in examining how different types of highlighting would influence people's social experience during a MT-mediated conversation. Although a small number of recent studies have discussed both shortcomings and benefits of using MT to support social interaction between multilingual dyads, they focused more on analyzing the content rather than improving the presentation of translated messages.

In studies conducted by Yamashita and colleagues [28][30], participants from China, Korea and Japan were asked to collaborate on a tangram task either with a shared second language (English) or with different native languages with MT support. The results indicated that MT hurt the collaboration by making it hard to establish common ground [8]. Other studies, however, suggest that MT can have a positive effect on social experience. In Hautasaari's [11] experiment, for example, participants from Finland and Japan collaborated on a shaping factory game. He found that using MT improved interaction in multilingual dyads by increasing the production of social-emotional messages.

In this study, we go beyond exploring the effect of MT itself by examining how highlighting keywords in a translated message influences social experience during MT-mediated communication. We hypothesized that highlighting keywords in each message would improve social experience during the communication because the highlighting would reduce cognitive load in information processing. Because communication with their partner was smoother and less effortful than with unhighlighted MT output, people would form more positive impressions of their partner and the quality of the collaboration.

H3: People will have more positive impressions of their partners when messages have keyword highlighting rather than no highlighting or random highlighting.

H4: People will have more positive impressions of their collaborations when messages have keyword highlighting rather than no highlighting or random highlighting.

As with comprehension, it is possible that highlighting random words in a message could potentially hurt the establishment of these positive social experiences relative to no highlighting by increasing communication difficulties. It could also have no effect. We thus posed the following two research questions:

RQ3: How will random highlighting affect impressions of a partner in comparison to messages with no highlighting?

RQ4: How will messages with random highlighting affect impressions of the success of the collaboration in comparison to messages with no highlighting?

METHOD

To test our hypotheses and examine our research questions, we conducted a laboratory experiment with a 3-level single factor (non- vs. random- vs. keyword highlighting on the message) within-subject design. Random highlighting was included as a second control condition, to ensure that the effects of keyword highlighting could be attributed to their value in making sense of the messages rather than to their ability to focus peoples' attention on a smaller number of words. All participants were native English speakers from the U.S. who had no knowledge of Mandarin Chinese. Each participant completed three ten-minute brainstorming tasks

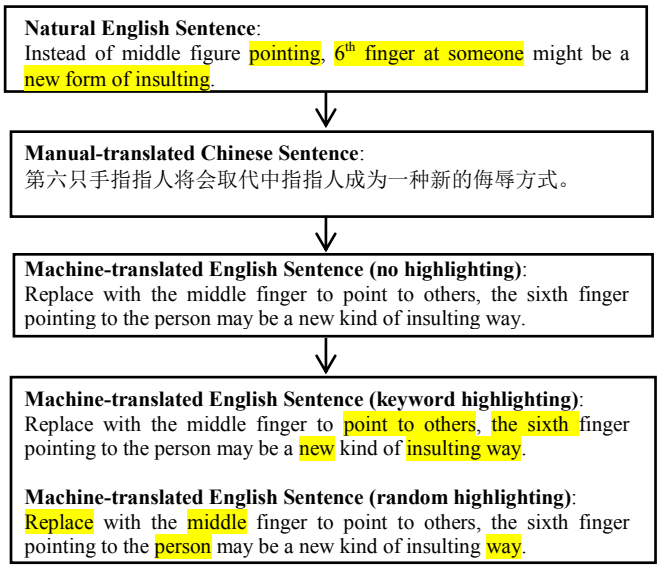

Figure 2. The procedure of generating the machine translated sentences with keywords highlighted (example from the *extra thumb* task).

with his/her Chinese-speaking partner (who was actually a confederate). After each task, the participant was asked about his/her communication experience.

Participants

Thirty-six undergraduate students (20 female) from a U.S. university participated in the study. All had lived in U.S. for more than 10 years and spoke English as their only native language. Their mean age was 20.25 years (*SD* = 1.66).

Materials

The purpose of this study was to examine how different rules for highlighting text in received messages influenced participants' performance and experience. However, since no current online chat tool or MT module can identify and highlight keywords of a sentence automatically, we developed materials specialized for this study so that the participant could receive instant messages with different types of highlighting during the brainstorming discussion.

Tasks. Three brainstorming tasks asked people to generate ideas about "what are the benefits and difficulties if people had *an extra thumb on each hand* / *a third eye on the back of head* / *two wings on the back* in the future" (see Figure 1). These tasks have been used in previous brainstorming studies, and the structure and difficulty of these tasks are considered equivalent [26][27]. During the task, the participant was asked to generate as many different ideas as possible with his/her partner about the given topic. All the brainstorming discussions in this study were text-based. In order to make the best use of time to generate ideas and to rule out alternative explanations for differences such as different amounts of social conversation, participants were asked to follow 2 rules: (1) don't talk about things which are irrelevant to the given topic with the partner (e.g., self-introduction, social chat, etc.), and (2) don't evaluate others' ideas.

Idea pool. An idea pool, from which sentences on the confederate's side were selected, was also developed. Since the participant was told to brainstorm with a Chinese-speaking partner, the messages received by the participant were supposed to be (1) machine-translated English sentences from the partner's Chinese sentences, and (2) with highlighting of particular words in the message depending on the experimental condition.

Thus, we needed to simulate a situation in which the ideas were generated by a Chinese speaker, then passed through machine translation. We started with a pool of ideas collected from a previous brainstorming study [27]. In that original dataset, there were 300 ideas in English, 100 ideas for each brainstorming task. Since our original idea pool was in English, we first asked two native Mandarin Chinese speakers to manually translate all the ideas into Mandarin Chinese. To assess the consistency of translations across translators, we counted the percentage of the sentences receiving the same translation (from English to Chinese) from two native Chinese speakers. During the first round of independent translation, 80% of the messages got the same Chinese translation from both translators. For the remaining 20% of the sentences, the translators discussed the inconsistencies to create a final translation of each message. Next, we used Google Translate to generate English translations of the 300 manually translated Chinese sentences to simulate the quality of translation that people would experience in MT-mediated communication.

Highlighting. Working independently, two native English speakers manually generated the keyword highlighting. Their goal was to identify words that captured the main points of each sentence. Since the Chinese to English machine translated sentences could sometimes be difficult to understand, the two English speakers started from the original English sentences obtained from [27]. They first highlighted the key words in those natural English sentences, then highlighted the counterparts of those keywords in the machine translated English sentences (Figure 2). The consistency between their keywords highlighting on the natural sentences was good (Kappa=0.78). They then discussed their highlighting until they came to agreement.

Random highlighting. The preparation of random highlighting was conducted directly on the machine translated English sentences. To keep the amount of highlighting roughly equivalent, for every idea we randomly selected the same number of words that had been highlighted in the keyword condition. Taking the idea shown in Figure 3, for example, we first counted how many chunks of keywords were highlighted in each idea under the keyword highlighting condition (4 chunks in this example), then randomly highlighted the same number of words in the same idea under the random highlighting condition. We randomized which words to be highlighted in this condition. This randomization was conducted for the purpose of getting a balanced distribution between messages with the entirely nonsensical highlighting and messages with the sporadically sensical highlighting.

Software and Equipment

Chat tool. We developed an online chat tool that could display messages with different types of highlighting (see Figure 3, next page). On the participant's side of the chat interface, the participant could type and receive messages similarly to common IM chat tools such as Gtalk (see the top of Figure 3). After logging in, participants could see if their partner had logged into the same conversation and was available to receive their messages. To input messages, participants typed sentences in the chat box and clicked the enter key.

Confederate interface. Because we wanted the same messages to appear in all conditions, confederates' communication was constrained to the idea pool described earlier. The messages in this pool were already highlighted

by color and categorized into no highlighting, random highlighting, and keyword highlighting. The confederate's side of the chat interface allowed him/her to select sentences from the idea pool.

Equipment. Both participants and the confederate were each seated at Dell computers with 25 inch monitors, separated by a divider. They wore headphones during the study so as not to be distracted by outside noise.

Procedure

Participants were brought to the laboratory and instructed about the rules to follow for the brainstorming tasks (no chat on irrelevant topics and no evaluation of one another's ideas). They then worked on three brainstorming tasks, one on each of the three brainstorming topics, and one with each of the three highlighting conditions. The order of tasks and highlighting conditions were counterbalanced across participants using a Latin Square design. Each task took 10 minutes.

Before each task, participants were told that they would have a brainstorming discussion with a partner who spoke Mandarin Chinese as his or her only native language. They were led to believe that they had a different partner for each task, though in reality there was only one partner, the confederate. We did this so that participants would rate their partners and the collaboration independently after each of the three highlighting conditions, rather than being influenced by earlier brainstorming discussions.

Participants typed their own ideas in their native language (that is, English) and received machine translated ideas (also in English but translated from Chinese) from their partners. Participants were made to believe that there was a MT module embedded in the chat tool for translation between English and Chinese.

During the conversation, the confederate selected machine translated English sentences from the pre-existing idea pool and sent these to the participant. The confederate configured the type of highlighting for ideas based on the current experimental condition. The confederate followed the same brainstorming rules as the participants, including no irrelevant chat during the discussion and no evaluation of the other person's ideas.

Measures

Participants answered the same questionnaire after each 10-minute brainstorming discussion. Responses to questions were provided on 7-point scales (1 = strongly disagree; 7= strongly agree). They were led to believe that they had a different Mandarin-speaking partner for each task.

Manipulation checks. Whether participants noticed the highlighted text was assessed by a single 7-point item, "I found some of the words in my partner's sentences were highlighted." Whether they found the highlighting sensible was assessed by a second 7-point item, "I found the highlighted words captured the key meaning of my partner's ideas accurately."

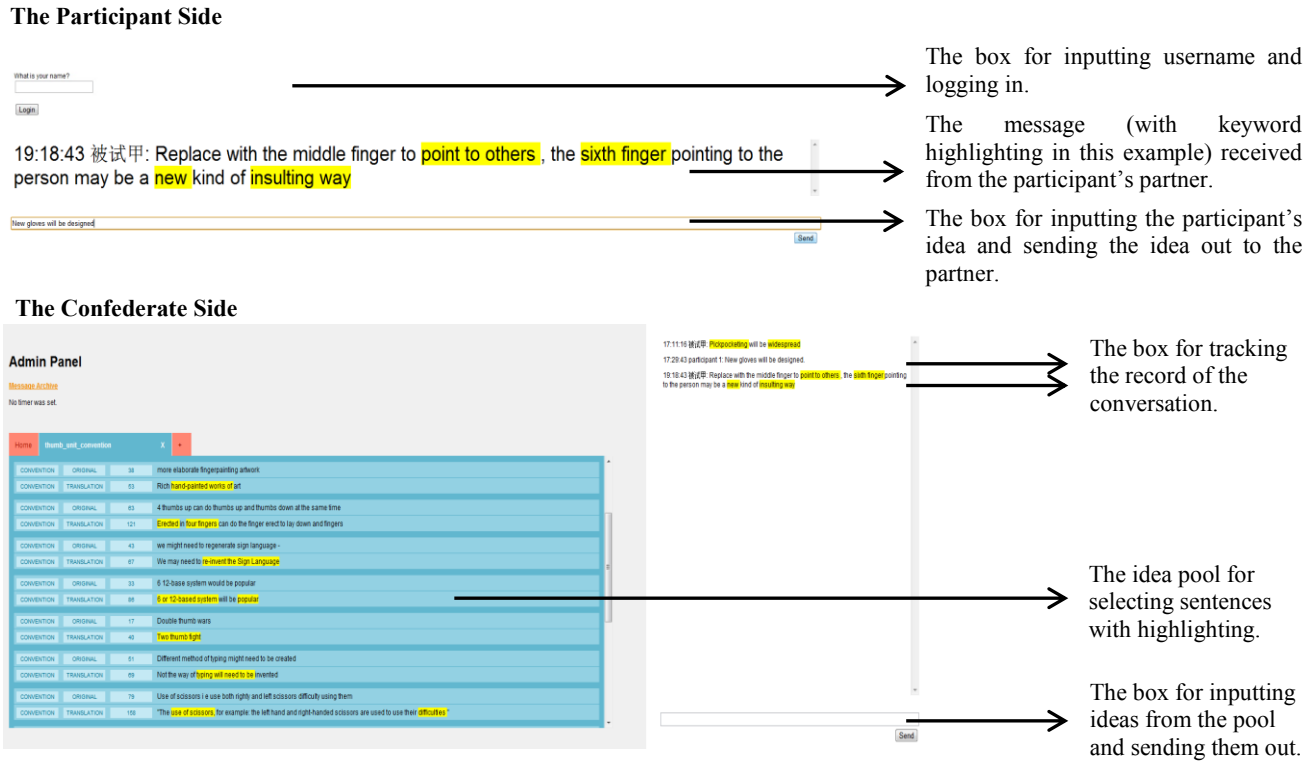

Figure 3. The interface of the chat tool developed in this study (top: participant side; bottom: confederate side).

Clarity of messages. People's understandings of the messages were assessed using a single item, "I felt my partner always expressed his/her idea clearly". Responses were on a scale of 1-7.

Distraction. The extent to which participants found the messages distracting was assessed using two questions ("The unclear information in my partner's ideas was distracting to me," "I had to think harder to understand the ambiguous information in my partner's ideas."). The two questions formed a reliable scale (Cronbach's α = .79) and responses were averaged to create a measure of distraction.

Impressions of partner. Participants' impressions of their partners were measured using seven 7-point Likert scales (e.g., "My overall impression of my partner was very positive", "My partner seems friendly"). Factor Analysis with Varimax rotation indicated that these seven questions loaded on a single dimension that accounted for 57% percent of the variance. Scores were averaged to create a measure reflecting the positivity of their impressions of their partner (Cronbach's α= .88).

Impressions of collaboration. Participants' impressions of the quality of their collaboration with their partner were measured using four 7-point Likert scales (e.g., "Generally, I'm satisfied with our collaboration on this task"). The questions formed a reliable scale (Cronbach's α= .73) and were averaged to create a measure of quality of collaboration.

Cognitive effects of highlighting. The cognitive effects of highlighting were measured using three 7-point Likert scales (e.g., "I paid more attention to the highlighted words rather than other non-highlighted words in the same sentence"). The questions formed a reliable scale (Cronbach's α= .95) and were averaged to create a measure of cognitive effects.

RESULTS

To test our hypotheses, we used repeated measures ANOVAs to detect whether participants' responses differed between highlighting conditions (non- vs. random- vs. keyword highlighting). Post-hoc LSD comparisons were used to compare performance between each pair of conditions.

Manipulation Checks

The manipulation checks indicated that participants noticed the highlighting ($F [2, 70] = 562.33$, $p < .0001$). Post-hoc comparisons showed that responses were higher for both keyword highlighting and random highlighting than no highlighting (both $p < .001$). Responses did not differ depending on the type of highlighting ($p = .90$). Participants rated keyword highlighting as significantly better at capturing the meaning of the message than random highlighting ($p < .001$). Therefore both manipulations were successful.

Understanding

H1 and H2 predicted that people's understanding of the messages would vary with different types of highlighting. Among all the three conditions of highlighting, the best understanding would be achieved under the keyword highlighting condition. To test this hypothesis, we examined the main effect of highlighting on ratings of message clarity (H1) and distraction (H2).

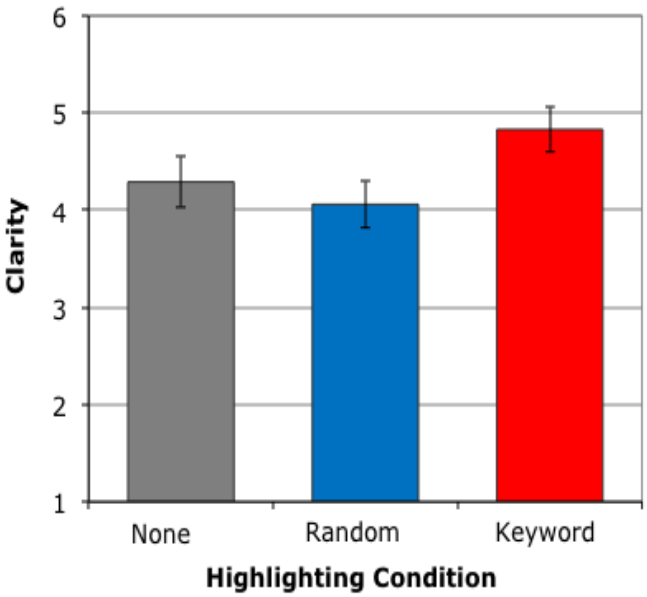

Figure 4. Mean clarity ratings by highlighting condition (error bars represent standard errors of the mean).

As shown in Figure 4, messages with keyword highlighting were better understood than messages with no highlighting or random highlighting ($F [2, 68] = 5.58$, $p = .006$). Post-hoc comparisons showed that clarity was significantly higher in the keywords highlighting ($M = 4.83$, $SD = 1.34$) than in the no highlighting condition ($M = 4.29$, $SD = 1.55$; $F [1, 34] = 5.59$, $p = .02$) or the random highlighting condition ($M = 4.06$, $SD = 1.41$; $p = .004$). These findings provide strong support for H1. RQ1 asked about the effects of random highlighting vs. no highlighting. Our analysis found no significant difference in clarity ratings between the two conditions ($p = .32$).

With respect to distraction, there was a significant main effect of highlighting ($F [2, 70] = 3.82$, $p = .03$) but the pattern of results was only partially consistent with H2 (see Figure 5). Ratings of distraction in the keyword highlighting condition ($M = 3.21$, $SD = 1.28$) were significantly lower than in the random highlighting condition ($M = 3.86$, $SD = 1.45$, $p = .01$) but did not differ significantly from those in the no highlighting condition ($M = 3.49$, $SD = 1.43$, $p = .20$). RQ2 asked about the effects of random vs. no highlighting on distraction ratings. We again found no significant difference in ratings between the two conditions ($p = .16$).

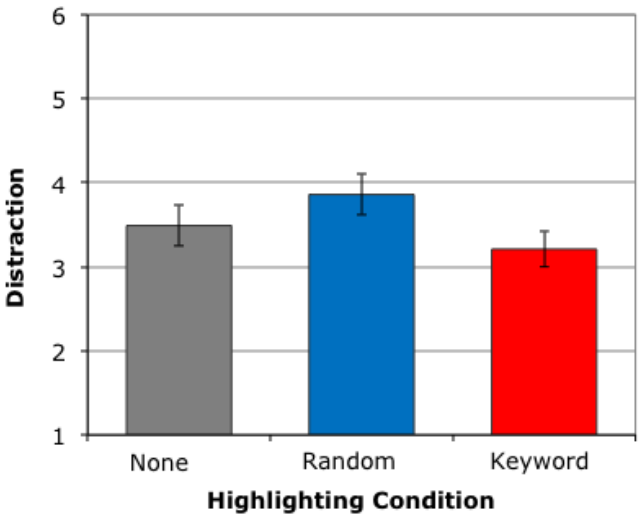

Figure 5. Mean distraction ratings by highlighting condition (error bars represent standard errors of the mean).

Social Experience

H3 and H4 predicted that people would rate their partners and their collaboration more favorably with keyword highlighting than with no highlighting or random highlighting.

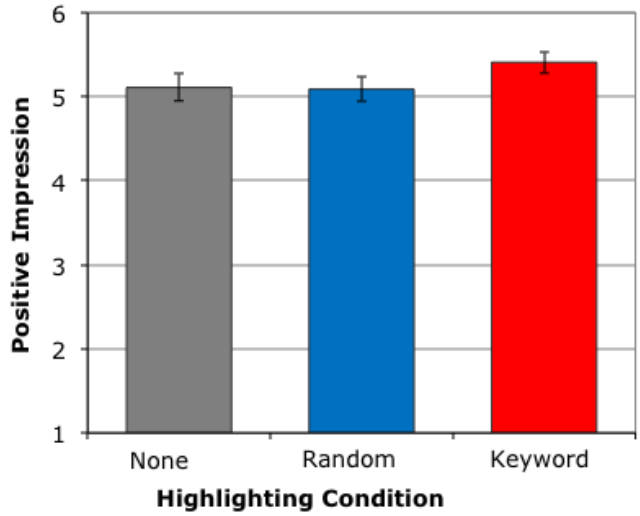

Figure 6. Mean impressions of partner by highlighting condition (error bars represent standard errors of the mean).

Consistent with H3, keyword highlighting led to more positive impressions of one's partner than random or no highlighting (F [2, 68] = 4.49, p = .02; see Figure 6). Impressions in the keywords-highlighting condition (M = 5.41, SD = 0.77) were significantly higher than in the no highlighting condition (M = 5.11, SD = 0.94, p = .01) or random highlighting condition (M = 5.09, SD = 0.89 p = .02). RQ3 asked about the effects of random highlighting vs. no highlighting on impressions of the partner. The two conditions did not differ significantly (p =. 87).

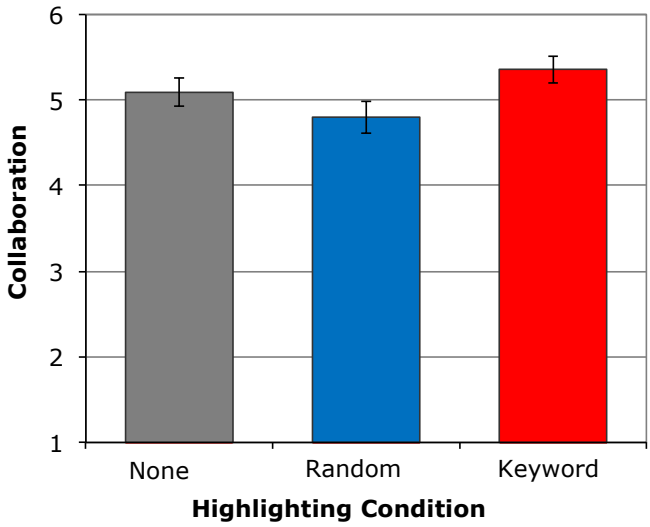

Figure 7. Mean ratings of the quality of the collaboration by highlighting condition (error bars represent standard errors of the mean).

As can be seen in Figure 7, participants' impressions of the success of their collaboration were also influenced by keyword highlighting (F (2, 68) = 6.13, p = .004), which fully supported H4. Ratings of how well the pair collaborated were significantly higher in the keyword highlighting condition (M = 5.36, SD = 0.92) than in the no highlighting condition (M = 5.09, SD = 0.90; p = .05) or random highlighting condition (M = 4.80, SD = 1.13; p = .003). RQ4 asked about the effects of random highlighting vs. no highlighting on perceptions of the collaboration. There was a tendency for ratings in the random highlighting condition to be lower than those in the no highlighting condition, but this difference was not significant (p = .10).

Effects of highlighting on cognitive processing

To assess whether ease of cognitive processing might explain the benefits of keyword highlighting, we examined people's responses to our measure of cognitive effects (which combined attention, understanding and speed of processing). The result indicated people processed information more effortlessly with keyword highlighting (M = 4.84, SD = 1.49) rather with random highlighting (M = 2.72, SD = 1.35; F [1, 35] = 59.54, p < .0001).

Correlation between understanding and social experience

We further tested the relationship between the understanding measures (clarity and distraction) and the social experience measures (impression of partner and quality of collaboration). As shown in Table 1, there were significant positive correlations between understanding and social experience in this study.

	1	2	3	4
1.Clarity	-			
2.Distraction	-.24*	.79**		
3.Impression	.58**	-.20*	.88**	
4.Collaboration	.64**	-.22*	.65**	.73**

*$p < .05$, **$p < .01$.

Table 1. Correlations between clarity, distraction, impression of the partner and quality of collaboration.

DISCUSSION

Our results suggest that highlighting can be used to facilitate MT-mediated communication. When keyword highlighting was provided, people understood messages better and had a more positive social experience.

The comparison between different types of highlighting with respect to understanding and social experience implied something surprising yet inspirational. As we hypothesized, keyword highlighting led to significant improvement in perceived clarity of messages and social experience. Furthermore, there was a positive relationship between message clarity and impressions of both a partner and the quality of collaboration. The correlation between clarity and quality of collaboration echoes Yamashita and colleagues' studies [28][30] by confirming the fundamental role of message understanding in MT-mediated collaboration. In their studies, the use of MT, as opposed to a shared second language, created confusion about the meaning of referring expressions and hindered group collaboration. Instead of trying to resolve the shortcomings of MT technology, our study focused on exploring an alternative way to support understanding. The results demonstrate the value of using simple designs to support comprehension. When message understanding is improved, we also improve the collaborative and social aspects of communication.

Further, the positive effect of using MT on social aspects of multilingual communication provides complementary knowledge for understanding the relationship between MT mediation and group collaboration. Although some previous studies have shown social benefits of using MT in multilingual communication [e.g., 11], this benefit was attributed to an increased use of social-emotional messages in MT communication. In our study, neither social chat nor idea evaluation was allowed (and we also checked the chat log to make sure participants followed this rule). In this case, the more positive impressions of the partner and of the collaboration can't be attributed to the use of social-emotional messages. Rather, it appears to be due to the ease of understanding a partner's messages. This provides an alternative way to think about how to improve social experience during MT-mediated communication.

Another interesting finding was the similarity of communication experience in the random highlighting and no highlighting conditions. Although random highlighting didn't provide any benefit, it didn't disrupt the communication either. It appears, based on their self-reports about their cognitive processing of the messages, that people didn't pay much attention to the highlighting if it was applied randomly. This suggests that automatic real time keyword highlighting may still be valuable even if the process is less than perfect. People may be able to overlook cases in which words that are not relevant to the meaning of the message are erroneously highlighted.

The mechanisms behind the keyword highlighting effect can also be inferred from people's self-reported cognitive processing in the keyword and random highlighting conditions. Highlighting appears to direct people's attention, but people also evaluate whether the highlighted words are worthy of their attention. The distribution of cognitive sources may be adjusted based on this process. Investigating the learning process in processing highlighted messages can be both of theoretical value to better understand human-information interaction and of practical value to the design of tools to support comprehension.

Design Implications

The findings have several design implications. First, the data indicate that keyword highlighting is an effective way to support MT-based communication at low cost. In the field of natural language processing (NLP), a huge body of work aims to improve algorithms for generating better translation between languages (e.g., [21]). However, the actual experience of using MT to communicate shows that current MT tools still have much space to improve for supporting real time communication.

From the psycholinguistics and communication literature, however, we learn that ambiguous and fragmentary information is a natural part of human language, while processing this imperfect information is part of people's everyday language use [6]. This suggests we could improve MT-mediated communication by not only focusing on the "MT" part, but also the "communication" part, such as developing enriched communication channels to support the processing of imperfect translation. The positive outcomes of simple keyword highlighting support this idea. The value of this communication-oriented approach is multi-faceted, including better understanding and social experience, and it is arguably at lower cost computation- and implementation-wise than traditional NLP work and other research which focuses on translation repair in HCI (e.g., [17][19]).

Further, the similar levels of clarity and experience in the random highlighting and no highlighting conditions suggest that people can tolerate semantically unimportant highlighting. Keyword highlighting may be implemented with simple, cost-effective algorithms, even if these sacrifice some precision in keyword identification. Even if the identification, and thus highlighting, of keywords is imprecise, it doesn't make a message harder to understand than one with no highlighting at all.

Technical Feasibility

The primary goal and contribution of our study was to explore whether working on technical solutions to translation interpretation, such as keyword highlighting, is a worthwhile use of energy by ensuring that highlighting has practical and cognitive value. Thus, we chose manual highlighting as a Wizard of Oz method to ensure high-quality highlighting. We countered this with the random condition to see whether poor-quality highlighting is harmful. Our results show that keyword highlighting is useful and efficient, even when the accuracy of highlighting is compromised. Such evidence makes it more likely that simple techniques can become plausible design solutions.

Our work opens up a design space for exploring technical solutions for tagging keywords in real time. When adding the feature of keyword highlighting to a MT tool, we need to further think about *who* will identify the keywords, *how* the agent performs the processing, and *what* methods can best highlight keywords under the time constraints of real-time communication. For example, it is possible to employ simple heuristics such as treating all verbs and nouns as keywords to obtain initial highlights first, then ask crowdsourcing workers to revise the results by clicking on individual words to add or remove highlights. When multiple workers perform this revision task at the same time, we can count the frequency of highlight-addition and highlight-removal to obtain a majority decision. Because the task that individual workers have to do is greatly simplified, a design of this sort that integrates machine processing and human processing and leverages crowd-based parallelism has the potential to perform quality and efficient highlighting.

Limitations and Future Directions

There were several limitations to our study that leave open important future work. First, the effect of keyword highlighting was gained by a simple form of highlighting, by presenting keywords with a yellow background. Other highlighting methods might weaken or strengthen this effect. In Wu and Yuan's [25] study, they compared the effects of highlighting with different colors on computer-based table content searching. Similar comparisons could be conducted in future studies to get deeper understandings on which visual form(s) of highlighting would be most helpful in supporting MT-mediated multilingual communication. The visualization could be extended to other formats besides color highlighting.

Further, the brainstorming task in our experiment was conducted in English with native English-speaking participants. It is not clear if the effects of keyword highlighting would be different when applied to other populations or to translation between different pairs of languages. People from East Asian cultures tend to distribute their attention to target and background objects more equally during information processing, whereas North Americans pay more attention to the target information rather than the background [20]. Perhaps this difference could weaken the effect of keyword highlighting during the MT-mediated communication among East Asians. Moreover, the culture-based communication style could also impact the understanding of messages in different languages. It will be interesting to see how such factors and the highlighting would interact to influence the quality of multilingual communication.

Further, the brainstorming rules could bring both benefits and limitations to our study. For the purpose of verifying the effects of keyword highlighting through a Wizard of Oz study, such rules were helpful for controlling the pool of sentences that confederates could choose from to present to participants. However, some threats to the generalizability of our findings were also raised from using such rules. As has been pointed out by Clark [7, 8], some conversational moves avoided in our task (e.g., asking questions) could play important roles in the establishment of common ground. To address this point, future studies should go further to examine the effect of keyword highlighting on MT-supported communication in less constrained settings.

CONCLUSION

Our research suggests that keyword highlighting is a useful way to improve the quality of MT-mediated communication. Compared with no highlighting and random highlighting, participants understood messages better and rated their partners and collaboration more highly when keywords in each machine translated message were highlighted. The findings inform the design of new tools to support communication and collaboration across language boundaries.

ACKNOWLEDGMENTS

This research was funded in part by National Science Foundation grant #1025425. The material is also based in part on work supported by the National Science Foundation, while Susan Fussell was working at the Foundation. Any opinion, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. We thank Leslie Setlock, Jean-Carlos Polanco, John Lee and Monica Chen for their assistance and the anonymous reviewers for their valuable comments.

REFERENCES

1. Abu-Hakima, S., McFarland, C., & Meech, J.F. (2001). An agent-based system for email highlighting. In *Proc. AGENTS 2001*, 224–225.
2. Aiken, M., Wang, J., Gu, L., & Paolillo, J. (2011). An exploratory study of how technology supports communication in multilingual groups. *International Journal of e-Collaboration* 7, 1, 17–29.
3. Barner-Rasmussen, W., & Bjorkman, I. (2007). Language fluency, socialization and inter-unit

relationships in Chinese and Finnish subsidiaries. *Management and Organizational Review, 3*, 105–128.

4. Bederson, B.B., Hu, C., & Resnik, P. (2010). Translation by iterative collaboration between monolingual users. In *Proc. Graphics Interface 2010*, 39–46.
5. Chessa, F., & Brelstaff, G. (2011). Going beyond Google Translate? In *Proc. the 9th ACM SIGCHI Italian Chapter International Conference on Computer-Human Interaction 2011*, 108–113.
6. Chi, E.H., Hong, H., L., Heiser, J., Card, S.K., & Gumbrecht, M. (2007). ScentIndex and ScentHighlights: productive reading techniques for conceptually reorganizing subject indexes and highlighting passages. *Information Visualization, 6*, 1, 32–47.
7. Clark, H. H. (1996). *Using Language*. Cambridge Press.
8. Clark, H.H., & Brennan, S.E. (1991). Grounding in Communication. In L.B. Resnick, R.M. Levine & S.D. Teasley (Eds.), *Perspectives on Socially Shared Cognition* (pp. 127–149). Washington, DC: APA.
9. Feely, A.-J., & Harzing, A.W.K. (2003) Language management in multinational companies. *Cross-Cultural Management: An Int'l Journal, 10*, 37–52.
10. Harzing, A-W., & Feely, A.J. (2008). The language barrier and its implications for HQ-subsidiary relationships. *Cross-cultural Management: An International Journal, 15*, 49–61.
11. Hautassari, A. (2010). Machine translation effects on group interaction: an intercultural collaboration experiment. In *Proc. ICIC 2010*, 69–78.
12. Henderson, J.K. (2005). Language diversity in international management teams. *International Studies of Management and Organization, 35*, 66–82.
13. Johnson, D., & Wen, S. (1976). Effects of correct and extraneous markings under time limits on reading comprehension. *Psych. in the Schools*, *13*, 454–456.
14. Kawasaki, K., Sasaki, H., & Yamaguchi, H. (2008). Effectiveness of highlighting as a prompt in text reading on a computer monitor. In *Proc. the 8th WSEAS Conference on Multimedia System and Signal Processing 2008*, 311–315.
15. Lim, J., & Yang, Y.P. (2008). Exploring computer-based multilingual negotiation support for English-Chinese dyads: Can we negotiate in our native languages? *Behaviour & Info. Tech.*, *27*, 139–151.
16. Marshall, C.C. (1997). Annotation: from paper books to digital library. In *Proc. the 2nd ACM international Conference on Digital Libraries 1997*, 131–140.
17. Miyabe, M., & Yoshino, T. (2010). Influence of detecting inaccurate message in real-time remote text-based communication via machine translation. In *Proc. ICIC 2010*, 59–67.
18. Miyabe, M., Yoshino, T., & Shigenobu, T. (2009). Effects of undertaking translation repair using back translation. In *Proc. IWIW 2009*, 33–40.
19. Morita, D., & Ishida, T. (2009). Collaborative translation by monolinguals with machine translators. In *Proc. IUI 2009*, 361–365.
20. Nisbett, R. (2003). *The Geography of Thought: How Asians and Westerners Think Differently and Why.* New York: Free Press.
21. Olive, J., Christianson, C., & McCary, J. (2011). *Handbook of Natural Language Processing and Machine Translation: DARPA Global Autonomous Language Exploitation*. Springer Press.
22. Olson, G.M., & Olson, J.S. (2000). Distance matters. *Human Computer Interaction, 15*, 139–178.
23. Tange, H., & Lauring, J. (2009). Language management and social interaction within the multilingual workplace. *Journal of Communication Management, 13*, 218–232.
24. Teevan, J., Dumais, S.T., & Liebling, D.J. (2010). A longitudinal study of how highlighting web content change affects people's web interaction. In *Proc. CHI 2010*, 1353–1356.
25. Wu, J., & Yuan, Y. (2003). Improving searching and reading performance: the effect of highlighting and text color coding. *Information & Management*, 40, 617–637.
26. Wang, H.C., Cosley, D., & Fussell, S.R. (2010). Idea Expander: supporting group brainstorming with conversationally triggered visual thinking stimuli. In *Proc. CSCW 2010*, 103–106.
27. Wang, H.C., Fussell, S.R., & Cosley, D. (2011). From diversity to creativity: stimulating group brainstorming with cultural differences and conversationally-retrieved picture. In *Proc. CSCW 2011*, 265–274.
28. Yamashita, N., & Ishida, T. (2006). Effects of machine translation on collaborative work. In *Proc. CSCW 2006*, 515–523.
29. Yamashita, N., & Ishida, T. (2006). Automatic prediction of misconceptions in multilingual computer-mediated communication. In *Proc. IUI 2006*, 62–69.
30. Yamashita, N., Inaba, R., Kuzuoka, H., & Ishida, T. (2009). Difficulties in establishing common ground in multiparty groups using machine translation. In *Proc. CHI 2009*, 679–688.
31. Yeh, C-L., & Chen, Y-C. (2004). Topic identification in Chinese based on centering model. *Proc. ACL 2004 Workshop on Reference Resolution and its Applications.*