Leveraging Machine Translation to Support Distributed Teamwork Between Language-Based Subgroups: The Effects of Automated Keyword Tagging

Yongle Zhang

College of Information Studies, University of Maryland, yongle@umd.edu

Dennis A. Owusu

Department of Computer Science, University of Maryland, dasamoah@umd.edu

Emily Gong

Department of Computer Science, University of Maryland, egong@terpmail.umd.edu

Shaan Chopra

College of Information Studies, University of Maryland, schopra7@umd.edu

Marine Carpuat

Department of Computer Science, University of Maryland, marine@umd.edu

Ge Gao

College of Information Studies, University of Maryland, gegao@umd.edu

Modern teamwork often happens between subgroups located in different countries. Members of the same subgroup prefer to communicate in their native language for efficiency, which increases the coordination cost between subgroups. The current study extends previous HCI literature that explores the effects of machine translation (MT) on crosslingual teamwork. We investigated whether automated keyword tagging would assist people's comprehension of imperfect MT outputs and, therefore, enhance the quality of communication between subgroups. We conducted an online experiment where twenty teams performed a collaborative task. Each team consisted of two native English speakers and two native Mandarin speakers. We provided MT support that enabled participants to read all subgroups' discussions in English before team meetings, but in two forms: with vs. without automated keyword tagging. We found MT with automated keyword tagging affected people's interaction with the translated materials, but it did not enhance translation comprehensibility in the context of teamwork.

CCS CONCEPTS • Human-centered computing • Human-computer interaction (HCI) • HCI design and evaluation methods

Additional Keywords and Phrases: Multilingual teamwork, Machine translation (MT), Automated keyword extraction

ACM Reference Format:

Yongle Zhang, Dennis A. Owusu, Emily Gong, Shaan Chopra, Marine Carpuat, and Ge Gao. 2021. Leveraging Machine Translation to Support Distributed Teamwork Between Language-Based Subgroups: The effects of automated keyword tagging. CHI Conference on Human Factors in Computing Systems Extended Abstracts (CHI '21 Extended Abstracts), May 8--13, 2021, Yokohama, Japan, 8 pages.

1 INTRODUCTION

Modern organizations often require subgroups located in different countries to achieve shared goals. Despite the common policy of using one required work language (i.e., English), members of the same subgroup often have local

discussions in their native languages [1, 2]. Prior work on language use and teams has outlined various benefits of this practice. For example, Ehrenreich pointed out that speaking a person's native language would maximize their efficiency and fluidity of communication [6]. A series of studies by Feely and Harzing also indicated that using a shared native language could strengthen the social bonding and trust within language-based subgroups [8]. In the context of teamwork, however, the aforementioned benefits often come at a cost at the team level. Neeley et al. found that there often lacked sufficient information exchange between language-based subgroups of the same team [24]. Other scholars further suggested that using different languages would result in social divisions between subgroups [21].

HCI researchers have explored ways to facilitate distributed teamwork across the language boundary. Many of these studies indicate the potential of leveraging machine translation (MT) to enable not only a flexible language choice of everyone, but also team-level communication [34]. However, they find that MT outputs could sometimes be incomprehensible for human communicants. Reasons leading to the low comprehensibility include semantic errors that distorted the meaning of the source sentences [10], inconsistent expressions for the same referent [33], and a lack of fluency as that of natural human language [4]. When these problems were salient, communicants would find it difficult to take advantage of MT for the joint task with others.

In the current study, we tested the idea of using automated keyword tagging to enhance people's comprehension of possibly imperfect MT outputs and, therefore, benefit team communication across the language boundary. This idea is backed up by findings from a few recent studies. For example, Green and colleagues examined how human translators conducted post-editing to improve MT outputs. They noticed that people tended to devote more attention to certain parts of the message (e.g., nouns and verbs) instead of others [11]. Gao and colleagues did a Wizard-of-Oz study where they asked human annotators to tag keywords in translated messages before using those messages in a mockup conversation. They found that participants perceived the tagged MT outputs to be more comprehensible than non-tagged ones [9]. Pan and Wang, similarly, showed that keywords annotated by crowd workers could facilitate crosslingual communication [25].

With the above findings as the proof-of-concept, we implemented automated keyword tagging to support teamwork between language-based subgroups. While auto-tagging presumably has a lower accuracy as opposed to human-annotated ones, the former has a significant advantage of generating keywords in real-time and without human effort. Our current study offered empirical evidence for understanding this cost-benefit tradeoff and its effects on teamwork. In particular, we found that the presence of keywords affected the way how participants interacted with the translated materials. It was also associated with a lower-level of workload experienced by participants. These findings inspire future solutions that leverage MT to support teamwork between language-based subgroups.

2 RELATED WORK AND RESEARCH QUESTIONS

2.1 Language-based subgroups in distributed teams

In distributed teams sitting across multiple countries, work communication often happens in "a cocktail of languages [2]." For example, Tange and Lauring interviewed employees at 14 Danish subunits of international companies. Interviewees reported that they chose Danish as the primary language to communicate when there were no meetings with English speaking teammates [28]. Hinds and colleagues observed that German speakers in English-speaking institutions used their native language to discuss work on a frequent basis [18]. A survey with employees at 70 different global corporations further showed that people often generated work related documents using the local language at each subunit of the corporation. They switched to English writing only in situations that they considered as necessary [27].

The diversity of language background draws a potential line dividing people into language-based subgroups. One consequence is that it hinders information sharing as well as social bonding between subgroups of the same team. In an ethnography study with one international cooperation, English speaking employees reported that their colleagues at other sites sometimes forwarded them emails containing conversations in the sender's local language. They felt lost and upset, especially when they were requested to respond to those emails [18]. In other studies, researchers found that members of each subgroup often treated the content of their subgroup discussions as taken-for-granted information for the rest of the team [5]. Monolingual teams have the option to enable a more transparent information exchange by asking all the subgroups to maintain a shared repository (e.g., an online space where people can forward their subgroup discussions for the entire team to view) [13]. The same practice, however, can turn out taxing when people need to take extra effort in translating their local information to the team's common language (i.e., English).

2.2 The promise and challenges of MT-mediated communication

The recent development of MT offers a low-cost way to bridge communication and information sharing across the language boundary. A small but growing number of HCI scholars have explored the idea of using MT to facilitate crosslingual teamwork. Yamashita and colleagues, for instance, studied MT-mediated team collaboration under various settings. Their findings showed that MT enabled speakers of different languages to complete joint tasks that were otherwise not achievable [34]. Wang and colleagues applied MT to assist brainstorming sessions between native speakers (NS) and non-native speakers (NNS) of English. They found that NNS generated a significantly higher number of ideas when they could communicate with the help of MT instead of using English as a second language [32].

Despite the values, MT sometimes deliver translation outputs that are hard to comprehend by human communicants. It opens the question of how people can make the best use of possibly imperfect MT outputs while minimizing the impact of translation errors. In response to this question, Gao and colleagues proposed that displaying multiple translation outputs at the same time could increase a person's confidence in interpreting the intended meaning of source messages [10]. Alternatively, Wang and colleagues suggested the idea of using image retrieval to complement a person's comprehension of written messages [31]. These solutions were proved successful through lab studies, but they both require human communicants to refer to additional information while processing the initial translation outputs. Thus, it is likely that the effect of these solutions would decrease as the volume of MT outputs increases.

2.3 Automated keyword tagging as a potential solution

Keywords can provide a compact representation of the message's semantic content without introducing additional information for people to process. Several HCI studies have used human-annotated keywords to assist MT-mediated conversations between speakers of different native languages. Findings from those studies showed that keyword tagging helped communicants focus on the intended meaning of MT outputs and overlook the errors [9]. However, the time cost and human effort in generating high-quality annotations are usually high.

Research in Text Mining, Information Retrieval and Natural Language Processing has offered automated ways to perform keyword tagging. It applies automated keyword extraction in determining which of the words occurring in the original text should be considered keywords, which is opposed to assigning keywords from a pre-defined taxonomy. Automated keyword extraction has been used to improve text summarization [35], text categorization [20], opinion mining [3], or document indexing [14]. Supervised extraction techniques rely on manual annotation to train feature-rich classifiers that categorize or rank keyword candidates [29, 30]. Unsupervised methods exploit co-occurrence patterns within a document to determine the importance of a potential keyword without any manual annotation [7, 23], and their

performance can rival that of supervised methods [17, 23]. Notably, keyword extraction remains a challenging task: as Hasan and Ng note "state-of-the-art performance on this task is still much lower than that on many core natural language processing tasks", and tagging precision is low (in the 20-30% range) on common benchmarks [17]. Most previous research in the technical field focuses on extracting keywords in written structured documents such as scientific articles or abstracts [23]. Much less attention has been paid to conversation logs, where topics might change more frequently and where documents are longer and have a less predictable structure [15]. In the current study, we adopt a keyword extraction strategy to balance accuracy and coverage (see section 3.3 for details).

2.4 Research questions

Building upon the above literature review, we investigated whether automated keyword tagging would assist people's comprehension of imperfect MT outputs and, therefore, enhance the quality of team communication across the language boundary. We conducted this investigation in a scenario that could be commonly observed in the real-world [5, 8, 18]: There were multiple language-based subgroups within one distributed team. Over the course of their joint work, each subgroup first had discussions in their native language, then proceeded to team meetings in English. We provided two forms of MT support to such teams: with vs. without automated keyword tagging. This support enabled participants to read all subgroups' discussions in English before team meetings. The research questions (RQ) we asked were:
RQ1. Does the automated keyword tagging affect people's comprehension of the translated subgroup discussions? How?
RQ2. Does the automated keyword tagging affect people's communication experience at English team meetings? How?

3 METHOD

3.1 Participants

We recruited 80 participants from one university. Half of the participants were NS of English who grew in the United States (17 female). Their mean age was 20.15 years (SD = 2.31). They reported having a medium level of experience in a crosslingual communication (M = 4.21, SD = 1.91 on a 7-point scale; 1 = never to 7 = very often). They were not very experienced in using translation tools or services (M = 3.00, SD = 1.18 on a 7-point scale; 1 = never to 7 = very often).

The rest of the participants were NNS of English who currently received education in the United States but grew up in China and spoke Mandarin as their native language (26 female). Their mean age was 23.72 years (SD = 3.46). These participants identified themselves as having a moderate level of English fluency (M = 4.93, SD = 0.83 on a 7-point scale; 1 = minimal fluency, 7 = native-level fluency). They reported having a medium level of experience in crosslingual communication (M = 4.60, SD = 1.11 on a 7-point scale; 1 = never to 7 = very often) and some experience in using translation tools or services (M = 4.27, SD = 1.38 on a 7-point scale; 1 = never to 7 = very often).

3.2 Task and Procedure

We adopted a modified online version of the Personnel Selection Task [26] that requires teams of quartets to complete. There are twenty teams in total. Each team consisted of two NS of English and two NS of Mandarin. The goal of the teamwork was to jointly evaluate four (pseudo) job candidates, then recommended one best candidate to take a research assistant position in the university. While members of the same team were all informed of the above task context, each participant received an exclusive set of information about those candidates (i.e., incomplete CVs that listing a subset of each candidate's prior research experience and work experience). The information received by NS of Mandarin was written in Mandarin, and that received by NS of English was written in English. This design mimicked the real-world

cases where members of distributed teams often held complementary information or expertise. It also increased the need for team members to exchange information across the language boundary.

All the participants attended the experiment through instant messaging (IM) and with pre-assigned pseudo and gender-neutral names. We developed an IM-based task platform which automatically saved participants' conversation logs and their mouse movements on the task interface over the entire course of the experiment. This task platform also allowed us to implement our experiment manipulations (i.e., MT, automated keyword tagging) when needed.

At the beginning of the experiment session, all the participants had subgroup conversations with their native speaking fellows. Participants were informed that they should exchange information with the other subgroup member using their shared native language. The subgroup conversation lasted for 15 minutes.

After that, participants received the text-based logs of the other subgroup's discussion. We used MT to translate Mandarin discussion logs for English speakers. We did not translate English discussions for Mandarin speakers because of two reasons. First, Mandarin speakers in this study held a moderate level of English fluency, which allowed them to comprehend English discussion logs. Second, previous research has shown that NNS in an English work environment were often in favor of asymmetric translations. They leverage translation support for producing ideas in their native language. However, they prefer receiving English messages as to how they are so that it avoids additional translation effort at a later point of the time [32]. We also manipulated the format of those discussion logs. Half of the subgroups received discussion logs with automated keyword tagging on them (Figure 1), whereas the other groups received discussion logs without tagging. Participants had 10 minutes to read the discussion logs.

Participant's (pseudo) name	Keywords tagged in the message	Message
Xiao	research experience	What kind of research experience did you have?
Lei	cross-, research experience, user	Research experience two, with cross-sectoral collaboration with members of the three other groups, creating user preference identification tools

Figure 1. A sampling excerpt of the discussion logs received by English speakers. In this excerpt, the first column indicates the pseudo name of each Mandarin speaker. The second column indicates automated keywords tagged in each message issued by its speaker. The third column indicates the complete (translated) version of the speaker's message.

As the last step of this task, participants proceeded to team meetings via IM. These team meetings all happened in English as a required common language among the quartets. Participants were informed that they should exchange information across subgroups so that the team could decide on one best candidate by the end of the meeting. The team meeting lasted for 15 minutes.

3.3 MT and Automated Keyword Tagging Tools

We used a state-of-the-art MT system using open-source tools and data to translate between English and Mandarin. We adopted a neural sequence-to-sequence model that was implemented in the AWS Sockeye toolkit[1]. The model design, training configuration, and training data are based on the top performing systems in public benchmarks. Specifically, we used an encoder-decoder model based on a 6-layer transformer network of size 512, with 8 attention heads, and a feedforward network size of 2048. The training data comprises 17.6 Million Mandarin sentences paired with their English translation, drawn from diverse news sources and United Nations corpora. The resulting system achieves a translation quality comparable with strong base transformer systems at the WMT2018 benchmark with a BLEU score of 23.6 on the

[1] https://github.com/awslabs/sockeye

official test set, and a qualitative analysis found that it produced translations that were comparable or more natural sounding that Google translate on text samples from this study.

We used a keyword tagging approach based on the dominant graph-based method. Specifically, we applied the TextRank algorithm [23]. TextRank[2] scores the importance of a word based on how it relates to other words in the given document. A word is considered important if it co-occurs with a large number of words, and with words that are important. To improve coverage, we also preset a small list of keywords that were manually curated for solving the task in our personnel selection materials. For example, "research experience" was manually set as a keyword to be tagged because it would intuitively matter in the candidate searching for a research assistant position. In addition to exact matches of the auto-extracted keywords, we allowed fuzzy matches for words with a small edit distance to the extracted keywords so that it would account for morphological inflections and typos. We validated the approach on a pilot dataset before the formal experiment, where it achieved a good balance of precision, recall and tagging speed. In the formal experiment, the system tagged 56 unique words and phrases. Six of them were from the human-curated list exclusively.

3.4 Measures

3.4.1 Perceived Comprehensibility of the Other Subgroup's Discussion

We measured participants' perceived comprehensibility of the other subgroup's discussion right after they read the discussion logs [22]. This measurement included four 7-point scales (e.g., "I understood what the other subgroup was saying," Cronbach's $\alpha = .83$). A higher average score indicated a better comprehensibility of the discussion logs.

3.4.2 Perceived Quality of Communication at Team Meetings

We measured participants' perceived quality of communication at team meetings right after they finished the discussion [22]. This measurement included five 7-point scales (e.g., "We clarified the meaning if there was a confusion of the messages exchanged at the team meeting," Cronbach's $\alpha = .80$). A higher average score indicated a better quality of communication at team meetings.

3.4.3 Perceived Workload of the Task

We measured participants' perceived workload of the task at the end of the experiment session [16]. This measurement included four 7-point scales (e.g., "How much time pressure did you feel due to the rate or pace at which the task or task elements occurred," Cronbach's $\alpha = .67$). A higher average score indicated a higher-level workload of the task.

3.4.4 Interaction Mode with the Other Subgroup's Discussion Logs

In addition to the above self-reports, we collected participants' mouse movements while they were reading the other subgroup's discussion logs. This measure indicated how a person interacted with the logs [12, 19]. Specifically, we counted the number of times when a person moved the scroll bar on their log reading interface to an opposite direction. Each person received a base number of 0 at the beginning of this calculation. We then added 1 to update this base number whenever the person moved their scroll bar towards a different direction from the previous one.

If a participant received a total number of 0 from this calculation, it meant that the person's log reading mode is entirely linear. If a participant received a total number above 0, it indicated that the person did some selective reading with the discussion log. The larger this number was, the more often the selective reading happened.

[2] https://github.com/DerwenAI/pytextrank

4 RESULTS

To explore our research questions, we conducted 2 (keyword availability: with vs. without automated keyword tagging) × 2 (language background: English vs. Mandarin) Mixed Model ANOVAs. Participants were nested within teams. The keyword availability and language background of participants were set as independent fixed variables. We set participants' demographic information (e.g., age, gender), previous experience in crosslingual communication, and previous experience in using translation tools or services as control variables in the models. However, the effects of those control variables were generally not significant. We presented the rest of the results with focuses on each independent variable's main effects as well as their interaction effects.

4.1 Perceived Comprehensibility of the Other Subgroup's Discussion

To answer RQ1, we conducted a 2 × 2 Mixed Model ANOVA on the perceived comprehensibility of the other subgroup's discussion. There was no significant main effect of keyword availability. There interaction effect between keyword availability and language background was not significant either. We found a significant main effect of language background: $F [1, 57.72] = 24.63$, $p < .001$.

Specifically, Mandarin speakers' perceived comprehensibility of the English discussion logs ($M = 5.43$, $SE = 0.18$) was significantly higher than English speakers' perceived comprehensibility of the translated Mandarin discussion logs ($M = 4.19$, $SE = 0.18$). The implementation of automated keyword tagging did not affect the result of this comparison.

4.2 Perceived Quality of Communication at Team Meetings

To answer RQ2, we conducted a 2 × 2 Mixed Model ANOVA on the perceived quality of communication at team meetings. No significant main effects or interaction effect were detected in this analysis. The perceived communication quality at team meetings did not vary according to our manipulations.

Figure 2. Mean workload (left) and mean selective reading (right) by keywork availably for Mandarin speakers and English speakers

4.3 Perceived Workload of the Task

We conducted a 2 × 2 Mixed Model ANOVA on the perceived workload, also in response to RQ2 (Figure 2). There were no significant main effects of keyword availability and language background. However, we detected a significant interaction effect between these two variables: $F [1, 58.89] = 4.18$, $p < .05$.

Specifically, English speakers perceived a lower level of workload when they have read a translated version of the other subgroup's Mandarin discussion logs with automated keyword tagging ($M = 3.13$, $SE = 0.22$) as opposed to without keyword tagging ($M = 3.77$, $SE = 0.23$): $F [1, 64.14] = 4.15$, $p < .05$. In contrast, Mandarin speakers rated their workload at an equal level no matter the English discussion logs they have read were with or without automated keyword tagging.

4.4 Interaction Mode with the Other Subgroup's Discussion Logs

We examined how participants interacted with the discussion logs under different conditions. We conducted a 2 × 2 Mixed Model ANOVA on the index of selective reading (Figure 2). There was no significant main effect of language background. There interaction effect between keyword availability and language background was not significant either. We found a significant main effect of keyword availability: $F [1, 17.81] = 11.73$, $p < .01$ (Figure 1).

Specifically, participants performed a greater amount of selective reading when they could read the other subgroup's discussion logs with automated keyword tagging ($M = 8.01$, $SE = 0.66$) as opposed to without keyword tagging ($M = 4.77$, $SE = 0.67$). The person's language background did not affect the result of this comparison.

5 DISCUSSION

In general, our results showed that the automated keyword tagging influenced certain aspects of the participants' task experience. English speakers reported a lower level of workload when they could take advantage of keywords to navigate the translated discussion logs. Our analysis of the interaction mode indicated a possible explanation for this result. Specifically, the keywords might offer English speakers convenient clues to recognize translated messages that shared the same main points. They would spend more efforts navigating the translated discussion logs when there were no keywords tagged in the logs. A similar pattern of increased selective reading was also revealed in Mandarin speakers' interaction with the discussion logs. However, we did not find a significant decrease in the workload perceived by those participants. The reason may lie in that the primary source of Mandarin speakers' workload comes from using English as a second language to perform the entire task. Having automated keyword tagging can guide Mandarin speakers to selectively read the English materials, but it does not change the task-level workload to a significant extent.

Interestingly, the current study did not show a significant association between keyword tagging and comprehensibility, although previous studies have found that human-annotated keywords would improve communicants' perceived comprehensibility of the message. We suspected that the above inconsistency might come from differences regarding the extent to which the task context was captured during the keyword generation process. While we tried to consider the context of a personnel selection task in the design of our automated keyword extraction method, human annotators may apply more sophisticated heuristics in identifying keywords that value the most in actual conversations toward a particular task goal. The incomplete capture of task context may also explain why participants in the current study experienced a less amount of workload after reading subgroup discussion logs that were with automated keyword tagging, but they did not find the keyword tagging helpful for improving task-oriented communication at team meetings. Further, previous studies often used human-annotated keywords to assist team communication at brainstorming sessions [9, 32]. The task context might matter less for message comprehension in those studies because participants were required to generate ideas that were independent from others' ideas. In our current task context, however, participants on the same team were expected to discuss a shared list of job candidates by building upon each other's thoughts. The comparison between our findings and those reported by previous studies indicated the importance of choosing and developing keyword tagging strategies against the particular context of communication.

ACKNOWLEDGMENTS

This research was funded in part by National Science Foundation grant #1947929 and the Maryland Catalyst Fund. We thank Weijia Xu and Sweta Agrawal for their input in designing the Mandarin-English Machine Translation system.

REFERENCES

[1] Nathalie Aichhorn and Jonas Puck. 2017. Bridging the language gap in multinational companies: language strategies and the notion of company-speak. Journal of World Business 52, 3: 386–403. https://doi.org/10.1016/j.jwb.2017.01.002

[2] Wilhelm Barner-Rasmussen and Christoffer Aarnio. 2011. Shifting the faultlines of language: a quantitative functional-level exploration of language use in MNC subsidiaries. Journal of World Business 46, 3: 288–295. https://doi.org/10.1016/j.jwb.2010.07.006

[3] Gábor Berend. 2011. Opinion expression mining by exploiting keyphrase extraction. In Proceedings of 5th International Joint Conference on Natural Language Processing, 1162–1170. Retrieved January 11, 2021 from https://www.aclweb.org/anthology/I11-1130

[4] Marine Carpuat and Mona Diab. 2010. Task-based evaluation of multiword expressions: a pilot study in statistical machine translation. In Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, 242–245. Retrieved January 11, 2021 from https://www.aclweb.org/anthology/N10-1029

[5] Catherine D. Cramton and Kara L. Orvis. 2003. Overcoming barriers to information sharing in virtual teams. Virtual Teams that Work: Creating Conditions for Virtual Team Effectiveness: 214–230.

[6] Susanne Ehrenreich. 2010. English as a business lingua franca in a German multinational corporation: meeting the challenge. The Journal of Business Communication (1973) 47, 4: 408–431. https://doi.org/10.1177/0021943610377303

[7] Samhaa R. El-Beltagy and Ahmed Rafea. 2010. KP-Miner: participation in SemEval-2. In Proceedings of the 5th International Workshop on Semantic Evaluation, 190–193. Retrieved January 11, 2021 from https://www.aclweb.org/anthology/S10-1041

[8] Alan Feely and Anne-Wil Harzing. 2003. Language management in multinational companies. Cross Cultural Management: An International Journal 10, 2: 37– 52. https://doi.org/10.1108/13527600310797586

[9] Ge Gao, Hao-Chuan Wang, Dan Cosley, and Susan R. Fussell. 2013. Same translation but different experience: the effects of highlighting on machine-translated conversations. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13), 449–458. https://doi.org/10.1145/2470654.2470719

[10] Ge Gao, Bin Xu, David C. Hau, Zheng Yao, Dan Cosley, and Susan R. Fussell. 2015. Two is better than one: improving multilingual collaboration by giving two machine translation outputs. In Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW '15), 852–863. https://doi.org/10.1145/2675133.2675197

[11] Spence Green, Jeffrey Heer, and Christopher D. Manning. 2013. The efficacy of human post-editing for language translation. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13), 439–448. https://doi.org/10.1145/2470654.2470718

[12] Qi Guo and Eugene Agichtein. 2010. Ready to buy or just browsing?: detecting web searcher goals from interaction data. In Proceeding of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval - SIGIR '10, 130. https://doi.org/10.1145/1835449.1835473

[13] Amar Gupta, Elisa Mattarelli, Satwik Seshasai, and Joseph Broschak. 2009. Use of collaborative technologies and knowledge sharing in co-located and distributed teams: towards the 24-h knowledge factory. The Journal of Strategic Information Systems 18, 3: 147–161. https://doi.org/10.1016/j.jsis.2009.07.001

[14] Carl Gutwin, Gordon Paynter, Ian Witten, Craig Nevill-Manning, and Eibe Frank. 1999. Improving browsing in digital libraries with keyphrase indexes. Decision Support Systems 27, 1–2: 81–104.

[15] Maryam Habibi and Andrei Popescu-Belis. 2013. Diverse keyword extraction from conversations. In Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), 651–657. Retrieved January 11, 2021 from https://www.aclweb.org/anthology/P13-2115

[16] Sandra G. Hart and Lowell E. Staveland. 1988. Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. In Advances in Psychology, Peter A. Hancock and Najmedin Meshkati (eds.). North-Holland, 139–183. https://doi.org/10.1016/S0166-4115(08)62386-9

[17] Kazi Saidul Hasan and Vincent Ng. 2014. Automatic keyphrase extraction: a survey of the state of the art. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 1262–1273. https://doi.org/10.3115/v1/P14-1119

[18] Pamela J Hinds, Tsedal B Neeley, and Catherine Durnell Cramton. 2014. Language as a lightning rod: power contests, emotion regulation, and subgroup dynamics in global teams. Journal of International Business Studies 45, 5: 536–561. https://doi.org/10.1057/jibs.2013.62

[19] Jeff Huang, Ryen W. White, and Susan Dumais. 2011. No clicks, no problem: using cursor movements to understand and improve search. In Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems - CHI '11, 1225. https://doi.org/10.1145/1978942.1979125

[20] Anette Hulth and Beáta B. Megyesi. 2006. A study on automatically extracted keywords in text categorization. In Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics, 537–544. https://doi.org/10.3115/1220175.1220243

[21] Dorte Lønsmann. 2014. Linguistic diversity in the international workplace: language ideologies and processes of exclusion. Multilingua 33, 1–2: 89–116. https://doi.org/10.1515/multi-2014-0005

[22] Leigh Anne Liu, Chei Hwee Chua, and Günter K. Stahl. 2010. Quality of communication experience: definition, measurement, and implications for intercultural negotiations. Journal of Applied Psychology 95, 3: 469–487. https://doi.org/10.1037/a0019094

[23] Rada Mihalcea and Paul Tarau. 2004. TextRank: bringing order into text. In Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing, 404–411. Retrieved January 11, 2021 from https://www.aclweb.org/anthology/W04-3252

[24] Tsedal Neeley, Pamela J. Hinds, and Catherine Durnell Cramton. 2009. Walking through jelly: language proficiency, emotions, and disrupted collaboration in global work. Social Science Research Network, Rochester, NY. https://doi.org/10.2139/ssrn.1414343

[25] Mei-Hua Pan and Hao-Chuan Wang. 2014. Enhancing machine translation with crowdsourced keyword highlighting. In Proceedings of the 5th ACM

International Conference on Collaboration Across Boundaries: Culture, Distance & Technology (CABS '14), 99–102. https://doi.org/10.1145/2631488.2634062

[26] Stefan Schulz-Hardt, Felix C. Brodbeck, Andreas Mojzisch, Rudolf Kerschreiter, and Dieter Frey. 2006. Group decision making in hidden profile situations: dissent as a facilitator for decision quality. Journal of Personality and Social Psychology 91, 6: 1080–1093.

[27] Esben S. Sorensen. 2005. Our corporate language is English: An exploratory survey of 70 DKsited corporations' use of English. Master Thesis in Language and Business Communication, Aarhus School Business, Aarhus.

[28] Hanne Tange and Jakob Lauring. 2009. Language management and social interaction within the multilingual workplace. Journal of Communication Management 13, 3: 218–232. https://doi.org/10.1108/13632540910976671

[29] Peter D. Turney. 2000. Learning algorithms for keyphrase extraction. Information Retrieval 2, 4: 303–336. https://doi.org/10.1023/A:1009976227802

[30] Chen Wang and Sujian Li. 2011. CoRankBayes: bayesian learning to rank under the co-training framework and its application in keyphrase extraction. In Proceedings of the 20th ACM International Conference on Information and Knowledge Management (CIKM '11), 2241–2244. https://doi.org/10.1145/2063576.2063936

[31] Hao-Chuan Wang, Dan Cosley, and Susan R. Fussell. 2010. Idea expander: supporting group brainstorming with conversationally triggered visual thinking stimuli. In Proceedings of the 2010 ACM Conference on Computer Supported Cooperative work (CSCW '10), 103–106. https://doi.org/10.1145/1718918.1718938

[32] Hao-Chuan Wang, Susan Fussell, and Dan Cosley. 2013. Machine translation vs. common language: effects on idea exchange in cross-lingual groups. In Proceedings of the 2013 Conference on Computer Supported Cooperative Work (CSCW '13), 935–944. https://doi.org/10.1145/2441776.2441882

[33] Naomi Yamashita, Rieko Inaba, Hideaki Kuzuoka, and Toru Ishida. 2009. Difficulties in establishing common ground in multiparty groups using machine translation. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09), 679–688. https://doi.org/10.1145/1518701.1518807

[34] Naomi Yamashita and Toru Ishida. 2006. Effects of machine translation on collaborative work. In Proceedings of the 2013 ACM Conference on Computer Supported Cooperative Work (CSCW '06), 515–524. https://doi.org/10.1145/1180875.1180955

[35] Yongzheng Zhang, Nur Zincir-Heywood, and Evangelos Milios. 2004. World wide web site summarization. Web Intelligence and Agent Systems 2, 1: 39–53.