Beyond Translation: Design and Evaluation of an Emotional and Contextual Knowledge Interface for Foreign Language Social Media Posts

Hajin Lim[1], Dan Cosley[1], Susan R. Fussell[1,2]
Department of Information Science[1], Department of Communication[2]
Cornell University
Ithaca NY 14850 USA
{hl934, drc44, sfussell}@cornell.edu

ABSTRACT

Although many social media sites now provide machine translation (MT) for foreign language posts, translation of a post may not suffice to support understanding of, and engagement with, that post. We present SenseTrans, a tool that provides emotional and contextual annotations generated by natural language analysis in addition to machine translation. We evaluated SenseTrans in a laboratory experiment in which native English speakers browsed five Facebook profiles in foreign languages. One group used the SenseTrans interface while the other group used MT alone. Participants using SenseTrans reported significantly greater understanding of the posts, and greater willingness to engage with the posts. However, no additional cognitive load was associated with using an interface that provided more information. These results provide promising support for the idea of using computational tools to annotate communication to support multilingual sense making and interaction on social media.

Author Keywords

Social Media; Machine translation; Multilingual communication; Cross-lingual communication; AI-augmented communication; AI-assisted communication; Sense making

ACM Classification Keywords

H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous;

INTRODUCTION

The user base of most popular social media platforms is multilingual [15, 24], and it is not uncommon for users to have contacts who sometimes post in languages they do not understand. For instance, a Facebook user who only understands English might have Chinese friends who sometimes post in Chinese. In this paper, we will refer to status updates in languages the user does not understand as "foreign language posts".

Publication rights licensed to ACM. ACM acknowledges that this contribution was authored or co-authored by an employee, contractor or affiliate of the United States government. As such, the United States Government retains a nonexclusive, royalty-free right to publish or reproduce this article, or to allow others to do so, for Government purposes only.
CHI 2018, April 21–26, 2018, Montreal, QC, Canada
© 2018 Copyright is held by the owner/author(s). Publication rights licensed to ACM.
ACM 978-1-4503-5620-6/18/04…$15.00
https://doi.org/10.1145/3173574.3173791

In principle, friends' foreign language posts might provide useful information about their context or emotional state, and people might want to interact with such posts by liking or commenting on them. Further, some recent studies show that having linguistically or culturally diversified social media connections can help people build cultural understanding and knowledge [28, 43].

However, language differences could pose a barrier to social media interactions across borders. Large scale analysis of Twitter data shows that differences in language and culture predict less communication volume between users [15] than between users who share the same language [16]. In addition, an eye-tracking study showed that monolingual English speakers paid less attention to and showed less willingness to engage with foreign language posts compared to English ones [30].

To promote better cross-lingual communication and interaction, popular social network sites such as Facebook, Twitter, and Instagram have begun to provide convenient machine translation (MT) tools for reading foreign language posts. However, a recent qualitative study found that simple translation features often do not suffice to help people make sense of foreign language posts, in part due to poor translation quality and in part due to difficulties understanding others' contexts and culture [29]. In addition, it revealed that people engage in a complex sense making process when encountering foreign language posts which involves collecting and combining cues and information from a variety of sources instead of relying exclusively on MT.

In this paper, we present a system, *SenseTrans*, that is designed to help people's sense making around foreign language posts. We see sense making as a useful lens for this situation because the need to understand something novel, unfamiliar, or confusing is a hallmark of sense making [31], which involves iterative processes of foraging and connecting information and generating multiple explanations [35]. To support such sense making processes about foreign language posts, we designed, developed and evaluated a system that provides not just a machine translation of the post text, but also contextual and emotional information derived from that translation. Contextual and emotional information is generally useful in grounding communication [5] and known to be important in cross-lingual sense making [29]. Using named entity recognition and sentiment analysis techniques from natural language processing (NLP), the system displays a *Keyword analysis* that provides contextual and cultural information about keywords in a post and an *Emotional analysis* that provides indicators of the overall polarity and presence of specific emotions in a post.

We examine the effects of this interface in a laboratory setting in which 24 native English speakers browsed five Facebook profiles and timelines in foreign languages constructed from real Facebook posts freely shared by informants. Half of the participants used SenseTrans, while the other group used MT alone. Participants using SenseTrans reported significantly greater understanding of the posts, but no additional cognitive load imposed by the extra information. Although these results need to be verified in real social media contexts, they suggest that augmenting machine translations with additional annotation is a promising approach to helping people connect better with friends who post in foreign languages.

In the remainder of the paper, we first discuss related work on social media and MT-mediated communication. We then present our design of SenseTrans, an emotional and contextual knowledge support interface. Next, we describe the procedure of a laboratory study that we conducted to evaluate SenseTrans and present our findings. We conclude by discussing the implications of this work for the design of tools that better support cross-lingual sense making and interaction on social media.

BACKGROUND

In this section, we first review literature on processes of communication on social media sites. Although these studies have generally focused on communication between speakers of the same native languages, they reveal important social processes that need support from tools designed to facilitate cross-lingual communication.

Communication on Social Media Sites

One key goal of social media use is staying connected with and aware of other people, particularly in social networking sites such as Facebook (e.g., [10,27]). Studies have identified a number of reasons why people use social network sites, including keeping updated on the activities of one's connections, building new relationships, and maintaining existing relationships (e.g., [7,26]). Zhao and Rosson emphasize the importance of informal communication on social media by consuming and engaging with others' status updates, in terms of both relational benefits (building common ground and sustaining a feeling of connectedness) and personal benefits (gaining valuable information or opportunities for one's interests and goals) [48].

Engaging in these social processes, mediated by consuming and interacting with others' posts, has been regarded as a relatively natural process that usually does not require extensive cognitive effort or sense making strategies [49]. However, when others post in a foreign language, these social processes could be hampered because of the cognitive effort and cultural distance experienced when interacting with text in a foreign language [29, 41].

For understanding foreign language posts, MT services are helpful, but not sufficient. Current MT quality is improving but still uneven, and struggles with user created content on social network sites that tends to be short, noisy, informal, colloquial, and relatively likely to contain idiosyncratic abbreviations, typos, and slang [6,8,17]. Further, a translation by itself may not help with key problems of understanding others' culture and context, which could harm people's ability to comprehend and engage with foreign language posts in social media [29].

MT-Supported Communication

Several technical and design approaches have been tried to better support text-based MT-mediated communication. Shigenobu proposed a back-translation system (translating through the partner's language back to one's own) that could help people monitor and repair MT outputs to improve communication [38]. Models that predict breakdowns have been developed to help partners detect and repair problems in MT mediated communication [46]. Also, Gao and her colleagues explored several design ideas such as showing two MT outputs at a time or highlighting key words in a translation, finding that such designs improved collaboration experience and message comprehension even with error-ful MT outputs [13,14].

At a high level, most of these design ideas aim at providing extra information or visual support to help people identify communication breakdowns and initiate repair to support grounding in real-time MT-mediated communication. A focus on repair is not so well suited for the asynchronous and often public communication that characterizes engagement with social media posts. However, we do see potential in the general idea of providing additional information about a post and its translation to support sense making around foreign language posts.

Particularly, we think providing contextual and emotional information along with translations can be useful for helping users make sense of foreign language posts. Among the knowledge types that could ground effective communication and collaboration, *emotional knowledge* about a partner's emotional states and *contextual knowledge* about the social and physical context within which communication takes place are particularly important [5]. These knowledge types also line up well with people's goals and strategies when making sense of foreign language

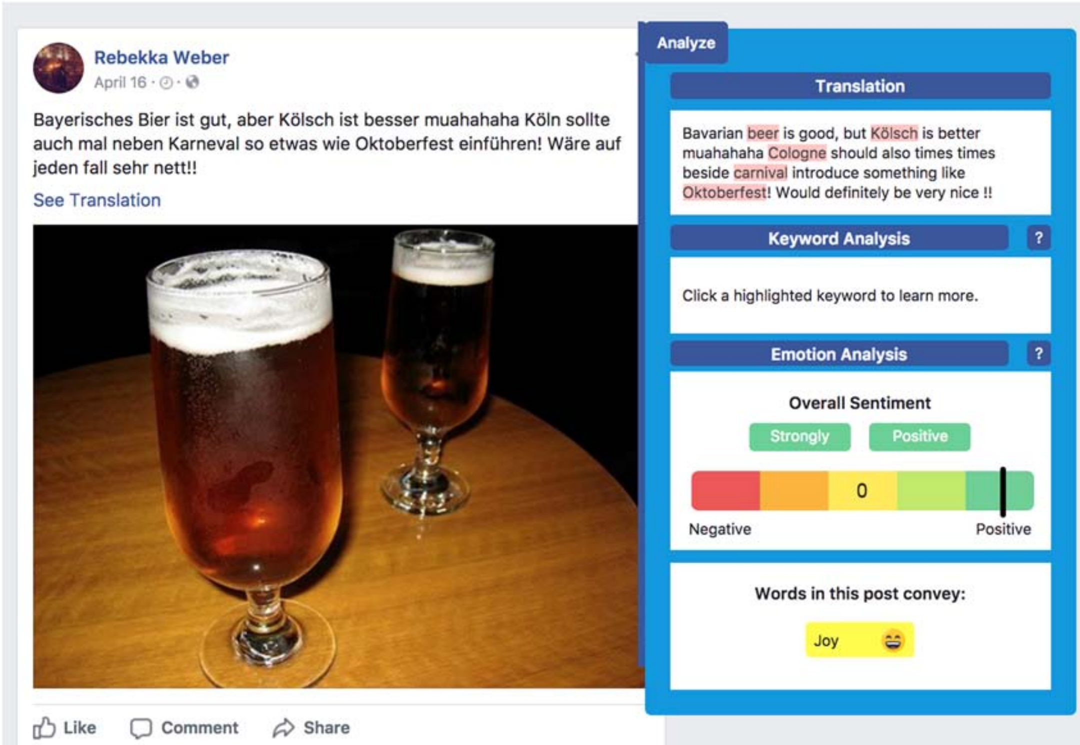

Figure 1. The interface of SenseTrans on a German Facebook post.

cultural meanings (e.g. what is Diwali? What is the historic background of this event?) [29]. Further, there are available NLP techniques such as sentiment mining (e.g., [3]) and named entity extraction (e.g., [1]) that can be used to generate information about the contextual and emotional content of a post.

SENSETRANS: AUGMENTING MT WITH CULTURAL AND EMOTIONAL ANNOTATIONS

Based on the considerations outlined in the previous section, we designed and developed *SenseTrans*, a tool to provide emotional and contextual information in addition to MT outputs to support better sense making of foreign language posts in social media. The SenseTrans prototype was developed as an Internet browser extension that works on Facebook newsfeed. While we considered different implementation options such as a new web service or mobile application, we concluded that it was more natural to integrate the prototype into people's normal context of browsing social media.

SenseTrans activates when users put their mouse cursor over a Facebook status update, at which time the system uses a Google language detection API [19] to decide whether the text is English or not. In the current version, we focus on native English speakers in part because of the greater availability of NLP tools in English [2,3] and in part because our accessible participant population was primarily native English speakers. However, users can disable it for specific languages other than English; for instance, a native Portuguese speaker might not want SenseTrans for Portuguese language posts.

If the detected language is non-English and not disabled by the user, SenseTrans shows an "Analyze" button that, when clicked, reveals three analysis buttons: *Translation*, *Keyword Analysis* and *Emotion Analysis*. The *Keyword analysis* and *Emotion analysis* buttons come with small "?" icons that briefly describe how the computations for each analysis work. Clicking the three analysis buttons reveals an *MT* space, a *Keyword analysis* space, and an *Emotion analysis* space, respectively. Figure 1 shows the interface after each of the three analysis buttons has been clicked, although the user has not yet clicked on a keyword highlighted in the translation.

In an earlier iteration of the system, clicking 'Analyze' immediately opened all three sections; however, informal testing raised concerns that showing all available information might be overwhelming. Therefore, we decided to allow people to open them separately so that they could access only the information they wanted, though at the cost of additional clicks.

How Contextual and Emotional Elements Were Computed and Represented

As shown in Figure 2, SenseTrans provides three kinds of information to support sense making: a machine translation of the text, a *keyword analysis* space that highlights and links important words in the translation to external resources to support cultural and contextual sense making, and an *emotion analysis* space that provides information about the overall sentiment and particular emotions present in the post.

Below, we discuss in more detail how each of the three analysis spaces works along with key decisions we made regarding their design. Our general design process was to try out several different APIs and UIs for generating and presenting the analyses. We conducted constant formative evaluations using different parameters for the APIs, and different interface layouts over different post contents. The research team and testers regularly evaluated the analyses against their own interpretations of posts in their native languages (Chinese, Korean, and German), and this guided our choices of specific tools. In general, we tried out and compared NLP tools that were stable and well-regarded in their domains and that support multiple languages, so that in the future the results can be generalized beyond English speakers.

In the *MT space*, users can see the machine translation result for the selected post. We wanted to include a translation because this is increasingly common in social media and a basic tool for making sense of foreign language text. After comparing available translation tools, we chose the Google Cloud Translate API [20] as having the best results and supporting multiple languages. As in most MT interfaces, the translated text is presented in a standard textbox.

The *Keyword analysis space* provides additional contextual information about a post by identifying named entities in the post and linking them to further information. We chose named entities extraction techniques for supporting cultural and contextual sense making because they often recognize culture-specific (e.g., food, historic site, holiday, custom) [33,34] and context-specific (e.g., current events, politicians, celebrities) information.

SenseTrans uses the TextRazor Named Entity Recognition API [40] to extract the list of important or proper nouns (such as names of people, organizations, places, and products) from the translated text and also from the original text for languages that TextRazor supports (e.g., Chinese, Dutch, Spanish, German). TextRazor provides a confidence score for each returned entity that ranges from 0.5 to 10. Through testing, we decided to exclude entities with a confidence score below 1 as they were often irrelevant. We also removed some types of entity that testers generally regarded as not useful for sense making about foreign language posts, notably numbers, day of the week, and URLs.

After generating the filtered list of entities, SenseTrans uses the Wikipedia API to display information about each of them. In general, the first one or two blurbs of the Wikipedia article are provided, along with a link to the Wikipedia page and any primary image associated with the article. When TextRazor returns a corresponding Wikipedia

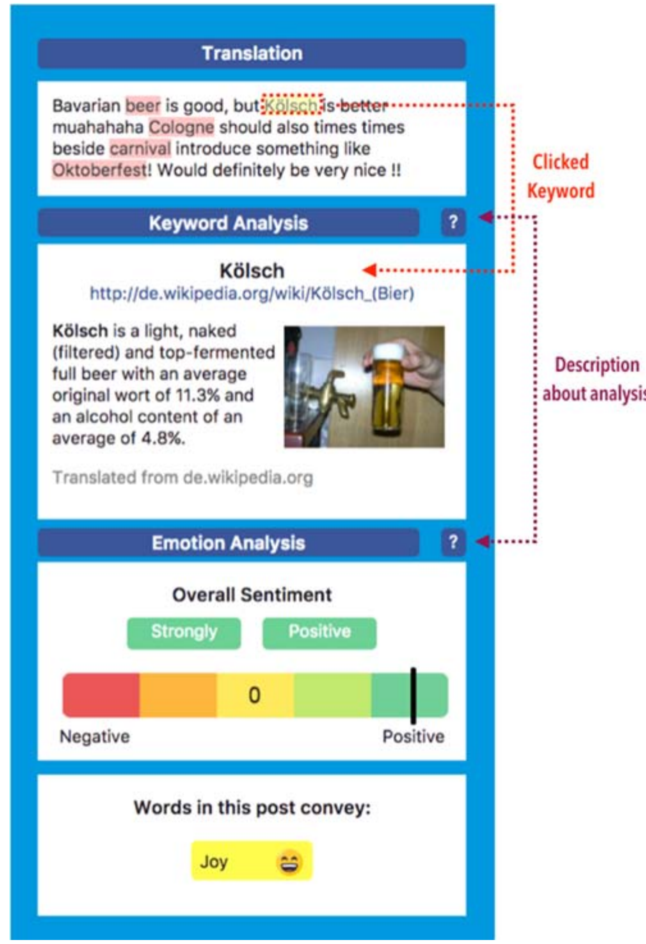

Figure 2. SenseTrans interface components (Top: MT Space, Middle: Keyword analysis space, Bottom: Emotion analysis space).

link for an entity, we make use of this link to provide information. When the API does not return a Wikipedia link for entities such as non-English terms or foreign figures, SenseTrans first tries to find a corresponding English-language Wikipedia page using the Wikipedia Search API; if one doesn't exist, it translates information from the matching Wikipedia page in the native language.

The *Emotion analysis space* provides information about the main sentiment and prevalent specific emotions in a post. We focused on these two analyses because studies have shown that people are less accurate at detecting emotional valence in foreign languages, which could harm cross lingual communication [22]. Providing better clues to other people's specific emotional states can also support socio-emotional communication [5].

SenseTrans uses the IBM Watson Natural Language Understanding API [25] to estimate both the overall sentiment and the presence of specific emotions in the text. As in the keyword analysis, the system uses the original text to calculate the sentiment and emotional scores if the

API supports the original language (e.g., French, German, Italian); English translations are used for posts written in unsupported languages.

The sentiment analysis algorithms return the numeric score of the document sentiment indicating the overall polarity of the post, with -1 representing highly negative and 1 representing highly positive overall sentiment. After a number of iterations, we decided to show the sentiment score as a marked bar with a text label such as “Moderately Positive”, rather than showing raw numbers such as 0.34, in order to help users get an overall sense of the sentiment without needing to interpret specific numbers.

Emotion detection analysis provides multi-dimensional sentiment analysis to interpret the types of feelings expressed in the words of the post. For each of five emotions (Joy, Fear, Sadness, Anger, Disgust), which are considered to be universal emotions [9]. For each emotion, the API returned a score ranging from 0% to 100% about whether the words in a post are likely to contain those emotions.

Our initial interface to these data showed a graph view labeled “emotion detection” that visualized each of the five scores. Informal pilot testing revealed several problems with showing raw values for all five emotions. As with overall sentiment scores, raw values can be hard to interpret—and users also worried that presenting raw values would lead people to think more about how the numbers were calculated rather than about making sense of the post. They saw both low, noisy raw values and cases where two or more values were relatively high as confusing. Computationally this is possible when posts contain multiple emotions [e.g., *Today was such a tough day* (Sad emotion words), *but I am so excited about the trip* (Joy emotion words)], but people reported wanting to focus on the dominant emotion.

Several design iterations based on this feedback led to the interface shown in Figure 2, which does not show the numeric results. Instead, it shows emojis for each emotion when at least one of the emotions was scored by the Watson API as over 50% likely to be present in the words of a post. We also replaced the label for the analysis from ‘Emotion detection’ to “Words in this post convey:” to make it clearer that the results were based on word-level analysis.

Testing MT, Emotion, and Keyword Analysis Quality

After arriving at the final version of the prototype, we performed a more systematic summative evaluation to validate that the translations and analyses might be useful results for sense making. We reasoned this would be most likely if the results corresponded with posters’ own evaluations about their status updates.

To assess this, we asked bilingual Amazon Mechanical Turk users to copy and paste their social media posts written in non-English languages and to rate the sentiment of their post on a 5-point scale from 1 = Negative to 5 = Positive. After we manually filtered out data that did not follow the instructions (e.g., some provided random text, or English posts), our dataset contained 241 social media posts written in 42 different languages. The top three languages were Spanish (20%), Hindi (17%), and Tamil (12%), providing a useful complement to the languages vetted by the research team during formative tests.

To evaluate the sentiment analysis scoring, we first calculated the Spearman correlation coefficient between the author’s rating (ranging from 1 = Negative to 5 = Positive) and the sentiment scores that we got from sentiment analysis API (ranging from -1 = Negative, 0 = Neutral, and 1 = Positive). This showed a moderate positive correlation between the author’s ratings and the API-generated sentiment score ($r = .40$, $p < .01$), suggesting that the sentiment algorithm provides a useful signal.

In addition, we ran precision and recall tests after categorizing the numeric values from both author’s ratings and sentiment scores into ‘Negative’, ‘Neutral’, and ‘Positive’ by setting up the thresholds. The overall accuracy of the result was 74% (Micro-average Precision: 61%, Recall: 61%, F-Score: 0.61).

To evaluate the accuracy and utility of the Named Entity extraction API, two undergraduate research assistants who were native English speakers read the translated English output for each post and marked entities they did not understand or wanted to know more about. We then checked the recall, examining how many of these entities had been returned by the Named Entity Identification API. Precision was less interesting for us because named entity extraction algorithms sometimes detect very general concepts such as “Germany”, “Beer”, and “Smart phone” and users could easily skip keywords that were not of interest or that were not very beneficial for cultural and contextual understandings. One coder marked 257 words, 210 of which were returned by the API (82% recall); the other marked 253, of which 220 were returned (87% recall). The micro-average recall was 84.3% and macro-average recall was 84.7%, which we judged as capturing most of the expressed information needs.

SYSTEM EVALUATION

We conducted a between-subject laboratory experiment to examine the effects of providing emotional and contextual information in addition to MT outputs on post comprehension and related measures. We chose a lab study with synthetic data rather than a field deployment using people’s own news feeds because our goal was to evaluate and validate the potential and design concept of SenseTrans. We aimed to explore whether using MT annotations to provide additional insight into other-language posts had potential for supporting sense making without (yet) having to account for factors such as history with an interaction partner or the prevalence or absence of foreign language posts in a person’s feed.

In the experiment, native English speakers were asked to browse five fictitious Facebook profile pages written in non-English languages (Spanish, German, Chinese, Korean, and Bahasa Malaysia) using one of two interface conditions: the full version of SenseTrans, or an MT-only version that only presented the translated text. After browsing each profile, participants answered a post-browsing survey about perceived comprehension of the status updates, willingness to engage with those posts, and social perceptions about the owner of profiles.

Research Questions and Hypotheses

Based on other studies showing that adding information or visual elements to translations could improve the perceived comprehension of messages [13,14], as well as findings about the need for emotional and contextual information in making sense of foreign language posts [29], we predict that using SenseTrans will improve perceived comprehension:

H1: People using SenseTrans will report higher perceived comprehension of foreign language posts than people using an MT-only interface.

This improved comprehension of foreign language posts could lead people to be more willing to engage with foreign language posts. Furthermore, it could increase people's willingness to be socially close with others, as annotating translations can also improve subjective impressions of the partner [14].

H2a: People using SenseTrans will be more willing to engage with individual status updates than people using an MT-only interface.

H2b: People using SenseTrans will be more willing to keep connection with the owner of profile pages than people using an MT-only interface.

Processing the emotional and keyword analyses in addition to the machine translation could increase cognitive effort; at the same time, this information could help people comprehend posts more easily, so it could lower overall cognitive effort. Therefore, we pose effects on cognitive load as a research question:

RQ1: How will SenseTrans affect participants' cognitive load compared to an MT-only condition?

Participants

We recruited 24 native English speakers (9 male), whose mean age was 19.4 years (*SD* = 1.2). To recruit participants, we posted flyers at various places around the campus of a large U.S. university and posted the study to an online participant recruitment platform at the university. On average, experimental sessions lasted 40-60 minutes. We compensated them either with \$10 or additional course credits. All participants were active Facebook users who used Facebook at least once a day. On average, our participants had 621 friends on Facebook (*SD* = 282) and they estimated that about one sixth (14.5%) of the status updates on their newsfeeds were written in a foreign language. All but one participant was aware of the "See Translation" feature for translating Facebook posts and all but two (91.6%) reported that they had translated their friends' Facebook posts using MT before. Except for one bilingual English and Spanish native speaker, participants had little or no fluency in languages other than English. Participants were randomly assigned to either the MT-only or the SenseTrans condition.

Study Materials

To construct the five fictitious profiles, we recruited 16 bilingual social media users whose first language was not English. We used a combination of word-of-mouth and snowball sampling to recruit bilingual social media users who were willing to share social media posts they had written in their native languages. We asked them to select real posts to share that did not include any personally identifiable information. Participants provided a total of 156 social media posts in eight different languages, and we compensated them with a \$5 gift card.

We used the collected data to create five different fictitious Facebook profiles, presented as belonging to college students from Germany, Venezuela, China, Korea, and Malaysia, choosing fake names for the profile owners based on the most common names in each region. Using a variety of source languages and profiles allowed us to diversify the cultural referents and reduce the chances that effects were based on a single language or set of content.

On each profile page, there were five status updates written in a non-English language based on the profile's identity (e.g. all status updates in the Korean student's profile page were written in Korean). We also tried to ensure that the posts in each profile represented common kinds of social media content, including typical topics of posts on a profile (e.g. travel, food, politics), as well as balancing the types of the post (personal vs. informative). We also added images to two of the status updates for each profile to make the profiles look more real.

Equipment

Participants were provided with a laptop computer (Intel Core i7 processor, 16 GB RAM) connected to a 24-inch LCD monitor and mouse for interacting with the Facebook profile pages and completing the surveys. Participants used the Google Chrome browser to view the profiles since SenseTrans and the MT-only version were both implemented as Chrome browser extensions.

Procedure

After signing a consent form, participants were told that the major goal of this study was to investigate how people consume and make sense of foreign language posts on Facebook. We told them that the owners of each profile they would browse were actual college students from all over the world who planned to come to U.S. as exchange students in the next semester.

We first asked participants to use the tool to browse publicly available Facebook sites such as 'HuffPost Arabi' (https://www.facebook.com/HuffPostArabi/) that are written in foreign languages, in order to let them become familiar with the SenseTrans or MT-only interface.

After they were comfortable with the tool, participants were then asked to browse each profile page and try to comprehend the status updates just as much as they wanted at their own pace. We asked them to use the browser extension to understand the posts instead of using the "See translation" feature on Facebook. Also, participants were asked to stay on the profile timeline page; clicking any link or moving to the other tabs of the profile pages (e.g. About, Photo) was not allowed. While they performed the tasks, we recorded their browsing activities (e.g., click, scrolling) to assess usage of specific tool features.

The five Facebook profiles were presented to participants in random orders. After browsed each profile pages, they filled out a post browsing survey about that profile and each status update on the profile. After they completed the survey, they moved to the next profile page and repeated this for all five profiles. After then, participants filled out a post task survey that asked about frequency of Facebook use and knowledge of the languages used in the profiles.

Measures

The post-browsing survey contained two parts, one about the whole profile and one that asked questions about each individual status update in the profile. The first part asked about overall comprehension, impressions of the profile owner, and cognitive workload. The second part showed each status update along with the translation and asked participants to rate their comprehension of and likelihood of engagement with each status update. Below we give details of each main measure.

Perceived comprehensibility. Participants were asked about their perceived comprehension of both the overall profile and of individual status updates. For profiles, participants were asked three 5-point Likert scale questions ("Overall, I am confident that I understand her/his profile page clearly", "I could understand what the profile owner tried to express clearly", and "I could understand the feelings of the profile owner clearly", from 1 = strongly disagree to 5 = strongly agree). The questions formed a reliable scale (Cronbach's α = .72) and were averaged to create a measure of "profile comprehension".

Similarly, to assess comprehension of each status update, we asked three questions using the same Likert scale ("It was easy to understand what this post was about", "I could understand what this post tried to express clearly", and "I could understand her/his feelings in this post clearly"). These formed a reliable scale (Cronbach's α = .80) and these were averaged to create a measure of "post comprehension".

Likelihood of Post Engagement. Each participant's likelihood of engaging with a single post was assessed by two questions ("I will click 'Like' for this post", "I will comment on this post") on 5-point Likert scales (1 = strongly disagree; 5 = strongly agree).

Willingness to keep connections with the owner of the profile. We used two measures of willingness to interact with a profile owner. The first was a social attraction measure based on McCroskey and McCain's interpersonal attraction scales [32]. Three items ("I think this person could be a friend of mine", "I would like to have a friendly chat with this person" and "S/he would be pleasant to be with") formed a reliable scale (Cronbach's α = .69), so they were averaged into a measure of "social attraction toward the owner of the profile". The second measure included two questions to examine participants' willingness to initiate friendship with the profile owner on Facebook ("I am willing to add the owner of this profile as my Facebook friends" and "I will accept the friend request from the owner of this profile"). These two questions formed a reliable scale (Cronbach's α = .81) and were averaged as a measure of "friendship intention".

Cognitive load. Perceived cognitive load in browsing each profile page was measured using three 5-point Likert scales adapted from the NASA Task Load Index [21] ('mental demand', 'effort' and 'frustration level', from 1 = low to 5 = high). The three items formed a reliable scale (Cronbach's α = .82) and they were averaged to create a measure of "cognitive workload".

RESULTS

In this section, we first examine how each group of participants (MT-only vs. SenseTrans) browsed Facebook profiles using the log data we collected. We then discuss the effects of annotating translations with emotional and contextual information on profile-level and post-level perceived comprehension (H1). Next, we report how SenseTrans affected participants' willingness to engage with a foreign language post (H2a) and to interact with the profile owners (H2b). Last, we examine whether SenseTrans affected cognitive effort in consuming foreign language posts (RQ1).

Participants using SenseTrans spent 142 seconds on average browsing each profile (28.4 sec per post), while those using the MT-only interface averaged 93 seconds on (18.6 sec per post). For most posts, participants using SenseTrans clicked all three of the Translation (94.8%), Emotion Analysis (85.9%), and Keyword Analysis (82.9%) buttons, though for the Keyword Analysis, participants didn't click on any keywords 40% of the time.

Comprehensibility

H1 predicted that participants using the SenseTrans interface will report higher perceived comprehension than participants using an MT-only interface. To test this hypothesis, we conducted two mixed model ANOVAs, one using profile comprehension and the other using post comprehension as the dependent measure.

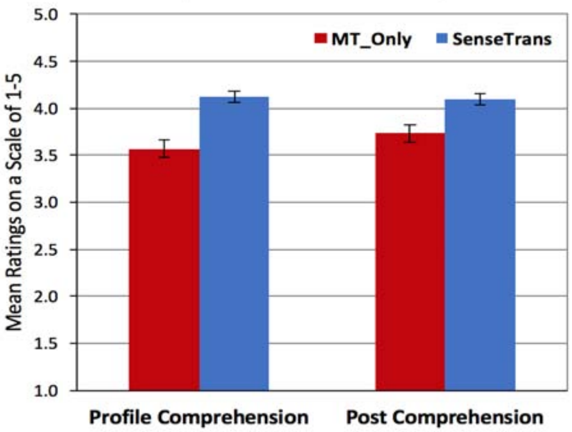

Figure 3. Self-reported profile comprehension and post comprehension by interface condition.

The results supported H1 at both the profile and post levels (Figure 3). At the profile level, we ran a mixed model ANOVA using participant as a random factor, interface condition as a fixed between-subjects factor, and profile language as a fixed within-subjects factor. The result showed a significant effect of interface condition on self-reported comprehension of the profile (F [1, 28.3] = 15.86, p < .01). Participants using SenseTrans had higher ratings for their comprehension levels (M = 4.12, SE = 0.1) than those using an MT-only interface (M = 3.57, SE = 0.1). Although comprehension for each profile language varied significantly (p < .01), there was no significant interaction between interface type and profile languages (p = .58).

At the post level, a mixed ANOVA using participant as a random factor, posts within profiles as repeated variable, and presence/absence of images as a fixed factor showed that using SenseTrans led to higher perceived comprehension ratings for each post (M = 4.1, SE = 0.06) than an MT-only interface (M = 3.74, SE = 0.06) (F [1, 129.8] = 19.86, p < .001). Participants perceived that they comprehended the posts with images better than posts without an image (F [1, 357.8] = 34.25, p < .001), but there was no significant interaction between interface type and the presence of post images (p = .69).

Willingness to Engage with a Post

H2a predicted that participants using SenseTrans would be more willing to engage with each status update by clicking “Like” or commenting on the post. To test this hypothesis, we used mixed model ANOVAs of the form described above.

We found a significant main effect of interface condition on both willingness to click “Like” and willingness to comment on a post (Figure 4). Participants using SenseTrans were more willing to “Like” posts (M = 3.01, SE = 0.14) than those using an MT-only interface (M = 2.34, SE = 0.14) (F [1, 66.3] = 10.98, p < .01). SenseTrans users were also significantly more willing to comment on posts (M = 1.92, SE = 0.12) than MT-only participants (M = 1.58, SE = 0.12) (F [1, 65.6] = 4.03, p < .05).

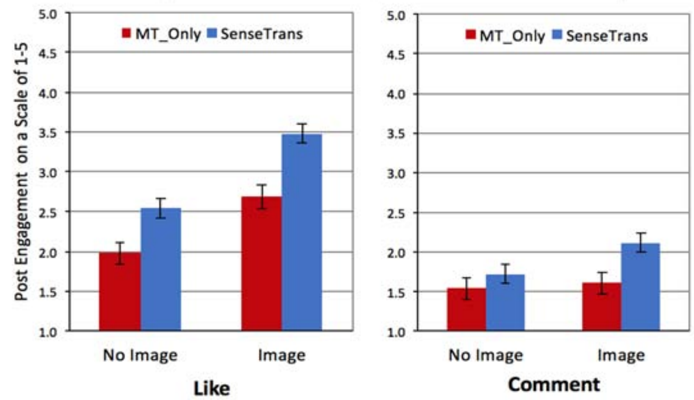

Figure 4. Willingness to engage with a post by interface condition (Left: Click ‘Like’, Right: Leave comment).

Consistent with previous work [30], people were more willing to like or comment on posts with images (p < .001 in both cases), but there was no significant interaction between interface type and the presence of post image on willingness to like a post (p = .22). However, there was a significant interaction between interface type and the presence of post image on the willingness to comment (F [1, 348.1] = 6.90, p < .01); participants who saw a post with images using SenseTrans were most willing to comment.

Willingness to Keep Connections with Profile Owners

H2b predicted that participants using SenseTrans will be more willing to keep connection with the owners of the profiles than participants using an MT-only interface. To test this hypothesis, we again constructed mixed model ANOVAs of the per-profile form described earlier, using the social attraction and friendship intention measures described earlier as dependent measures. The results (Figure 5) did not support H2b for either social attraction (F [1, 31.4] = 0.18, p = .68) or the intention to initiate friendship (F [1, 26.3] = 0.54, p = .47).

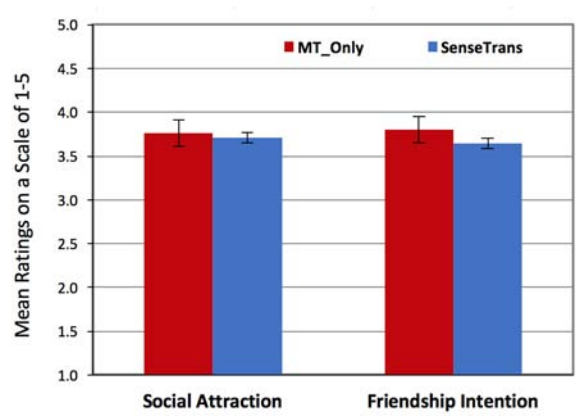

Figure 5. Social attraction and friendship intention for the profile owner by interface condition.

Cognitive Workload

Finally, RQ1 asked how the additional information provided in SenseTrans would affect cognitive workload around comprehending Facebook profiles. We conducted mixed model ANOVAs of the per-profile form described earlier with cognitive workload as the dependent measure. The results showed no significant difference in cognitive effort between the SenseTrans condition ($M = 2.08$, $SE = 0.15$) and the MT-only condition ($M = 2.29$, $SE = 0.15$) ($F[1, 25.51] = 1.05$, $p = .314$). See Figure 6.

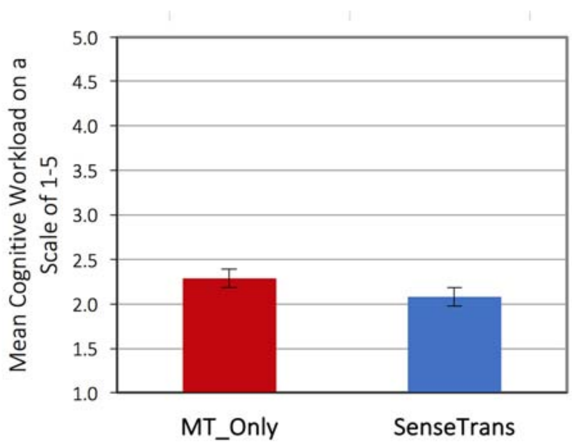

Figure 6. Cognitive workload by interface condition.

DISCUSSION

Overall, our results suggest that providing additional emotional and contextual information along with machine translations can improve perceived comprehension of foreign language social media posts. Furthermore, it might encourage more interaction across language boundaries in social media. Below, we discuss these results in more detail and outline further ideas around ways to use computational techniques to support sense making and cross-lingual communication on social media.

Consistent with H1, we found that participants reported that they comprehended foreign social media posts better when a machine translation was annotated with emotional and contextual information. Further, the additional information imposed no additional cognitive load for people who used SenseTrans despite their spending over 50% more time browsing profiles than people who used the MT-only version.

Because our focus was on users' self-evaluation of their understanding of the posts, we cannot conclude that SenseTrans led to more accurate and objective understanding of the content, but improving perceptions of understanding in cross-lingual sensemaking could have value in and of itself.

For instance, participants using SenseTrans were more willing to engage with foreign language status updates, consistent with H2a. We suspect that this is because improved perceptions of one's understanding of a foreign language post might lead to greater feelings of self-efficacy around responding appropriately to it. Greater self-efficacy should lead people to be more motivated and feel more empowered [37,47] to engage with such posts—in turn helping realize and utilize the cross-lingual and cross-cultural potential of social networks sites like Facebook.

Greater self-efficacy would be a problem, however, if perceived comprehension did not align well with actual comprehension. This is a tricky question in social media, since there are often gaps between posters' intentions and audience understandings even within a single language community [36]. Our summative testing of the analyses against people's own perceptions of their posts and their information needs in foreign language poses suggests that SenseTrans probably does support actual as well as perceived comprehension. Still, future studies should examine its effects on both actual and perceived comprehension, both for immediate goals of supporting specific interactions and longer-term benefits around exposure to and knowledge of cultural diversity.

Contrary to H2b, the increased comprehension of and willingness to engage with individual posts offered by SenseTrans did not translate into an increased desire to connect with profile owners. This is somewhat inconsistent with prior findings from studies of augmented MT interfaces, where the additional information improved perceptions of the interaction partner [13,14].

Unlike those studies, in which people chatted directly through MT, here there was no direct interaction between participants and the owner of the profile that might have served to build impressions and social perceptions. It is also possible that since this was a lab study, participants might not have seen any meaningful way to interact with profile owners (though we did try hard to create realistic profiles and posts). We made this realism tradeoff to ensure that people encountered a variety of foreign language content and to increase experimental control, but future work needs to evaluate how these tools would affect people's actual social interactions over time in a real deployment.

Another area of future work would be to tease out the separate contributions of the emotional and contextual knowledge elements to these results. We considered testing these components separately in the current study but decided not to because we wanted to see if the general approach of annotated MT had potential, and felt that deploying the full interface was most valuable for this. Usage statistics suggest people accessed the emotional information more often than contextual information, but that is only a first cut; future studies should look at how the individual components and other types of annotations affect one's comprehension and willingness to engage with posts.

Another limitation is that, although it made sense to focus on native English speakers given the state of NLP tools and our participant population, we do not know how native speakers of other languages would react to these tools.

Therefore, future work will be necessary to evaluate versions of these tools aimed at native speakers of other languages.

Design Implications and Future Directions

In addition to the straightforward implication that providing emotional and contextual information in MT systems can be helpful, our results suggest other ideas for tools that better support MT-mediated sense making in social media.

One interesting direction would be to explore how to augment and use images with emotional and contextual information. In the current study, we focused on translations and text, in part because the tools for manipulating them are more readily available. However, images contain much cultural and contextual information that could support such sense making [29], and study participants here were more confident in both their comprehension of and willingness to interact with status updates that included pictures. One path to effective use of images to support cross-lingual communication would be to use information retrieval techniques to use images themselves as ways to augment text, parallel to other work that uses image as a parallel channel to text for brainstorming [42] and second-language reading [18]. Another direction would be to use computer vision techniques such as object recognition, image classification, and face and gesture identification to extract, highlight, and annotate information about important cultural and emotional elements in the images people post [e.g. 12,44].

More generally, tool designers might consider of providing emotional and contextual information alongside machine translations in settings other than social media. We focused on sense making around social media posts (status updates, photos, and the like) since our main goal was improving sense making and encouraging cross-language interaction in social media. However, making sense of cultural context and emotional state is just as important in synchronous communication. Also, in longer-form of communication contexts such as composing and interpreting emails and news articles, our design approach could be adapted to these situations as well.

We also suspect that this kind of sense making support could help people even when they interact with people and content in languages they are fluent in; as argued in [36], understanding gaps are common even within a single language community. For native speakers, adding additional annotations might be able to help them grasp the main idea of social media posts even more quickly and effectively. For second language learners, it could be helpful to understand the cultural and contextual meaning of a post beyond its literal interpretation.

CONCLUSION

We present SenseTrans, a tool that provides emotional and contextual information generated by natural language analysis in addition to machine translation to support people's sense making process for foreign language social media posts. We evaluated SenseTrans in a laboratory experiment in which native English speakers browsed five Facebook profiles in foreign languages. Participants using SenseTrans reported significantly greater understanding of the posts and more willingness to engage with the posts, but no additional cognitive load.cThese results provide promising support for the idea of using NLP and other computational tools to annotate communication to help people better understand and interact with others in social media across language barriers, and potentially in many other contexts as well.

ACKNOWLEDGEMENTS

This work was supported by National Science Foundation #1421929 and #1318899. The paper was written while Dan Cosley was serving at NSF; it presents the views of the authors and does not necessarily reflect the views of NSF. We thank Karrie Shi, Joyelle Gilbert and Sangmin Kim for helping develop the prototypes and Masha Rosca and Vicky Chou for helping collect and analyze data.

REFERENCES

1. Rami Al-Rfou, Vivek Kulkarni, Bryan Perozzi, and Steven Skiena. 2015. Polyglot-NER: Massive multilingual named entity recognition. In *Proceedings of the 2015 SIAM International Conference on Data Mining*, 586–594.
2. Matheus Araujo, Julio Reis, Adriano Pereira, and Fabricio Benevenuto. 2016. An evaluation of machine translation for multilingual sentence-level sentiment analysis. In *Proceedings of the 31st Annual ACM Symposium on Applied Computing*, 1140–1145
3. Alexandra Balahur and Marco Turchi. 2014. Comparative experiments using supervised learning and machine translation for multilingual sentiment analysis. *Computer Speech & Language* 28, 1: 56–75.
4. Mikhail Bautin, Lohit Vijayarenu, and Steven Skiena. 2008. International Sentiment Analysis for News and Blogs. In *Proceedings of ICWSM*.
5. Charles R. Berger. 2011. Knowledge structures and social interaction. In *the SAGE handbook of interpersonal communication*, Mark L. Knapp, and John A. Daly (eds). SAGE Publications
6. Kalina Bontcheva and Dominic Rout. 2014. Making sense of social media streams through semantics: a survey. *Semantic Web* 5, 5: 373–403.
7. Moira Burke, Robert Kraut, and Cameron Marlow. 2011. Social capital on facebook: differentiating uses and users. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '11). ACM, New York, NY, USA, 571-580. DOI: https://doi.org/10.1145/1978942.1979023

8. Jordi Carrera, Olga Beregovaya, and Alex Yanishevsky. 2009. Machine translation for cross-language social media.
9. Paul Ekman. An argument for basic emotions. 1992. *Cognition & emotion* 6, 3-4: 169-200.
10. Nicole B. Ellison, Charles Steinfield, and Cliff Lampe. 2011. Connection strategies: Social capital implications of Facebook-enabled communication practices. *New media & society* 13, 6: 873-892.
11. Nicole B. Ellison, D Y.vette Wohn, and Christine M. Greenhow. 2014. Adolescents' visions of their future careers, educational plans, and life pathways The role of bridging and bonding social capital experiences. *Journal of Social and Personal Relationships* 31, 4: 516–534.
12. Ali Farhadi, Mohsen Hejrati, Mohammad Amin Sadeghi, Peter Young, Cyrus Rashtchian, Julia Hockenmaier, and David Forsyth. 2010. Every picture tells a story: generating sentences from images. In *Proceedings of the 11th European conference on Computer vision*: Part IV (ECCV'10), Kostas Daniilidis, Petros Maragos, and Nikos Paragios (Eds.). Springer-Verlag, Berlin, Heidelberg, 15-29.
13. 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*, 449–458.
14. 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*, 852–863.
15. Ruth García-Gavilanes, Yelena Mejova, and Daniele Quercia. 2014. Twitter ain't without frontiers: economic, social, and cultural boundaries in international communication. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*, 1511–1522.
16. Scott A. Hale. 2014. Global connectivity and multilinguals in the Twitter network. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 833–842.
17. Bo Han, Paul Cook, and Timothy Baldwin. 2013. Lexical normalization for social media text. *ACM Transactions on Intelligent Systems and Technology (TIST)* 4, 1: 5.
18. Cheng-Hsien Han, Chi-Lan Yang, and Hao-Chuan Wang. 2014. Supporting second language reading with picture note-taking. In *CHI '14 Extended Abstracts on Human Factors in Computing Systems* (CHI EA '14). ACM, New York, NY, USA, 2245-2250. DOI: https://doi.org/10.1145/2559206.2581204
19. Google. Detecting Languages | Google Cloud Translation API Documentation | Google Cloud Platform. Retrieved September 18, 2017 from https://cloud.google.com/translate/docs/detecting-language
20. Google. Cloud Translation API - Dynamic Translation | Google Cloud Platform. Retrieved September 17, 2017 from https://cloud.google.com/translate/
21. Sandra G. Hart and Lowell E. Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in psychology* 52: 139–183.
22. Ari M J Hautasaari, Naomi Yamashita, and Ge Gao. 2014. Maybe it was a joke: emotion detection in text-only communication by non-native english speakers. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*, 3715–3724.
23. Rebecca A. Hayes, Caleb T. Carr, and Donghee Y. Wohn. 2016. One click, many meanings: Interpreting paralinguistic digital affordances in social media. *Journal of Broadcasting & Electronic Media* 60, 1: 171–187
24. Lichan Hong, Gregorio Convertino, and Ed H Chi. 2011. Language Matters in Twitter : A Large Scale Study. *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*, 1: 518–521.
25. IBM. Watson Natural Language Understanding. Retrieved September 17, 2017 from https://www.ibm.com/watson/services/natural-language-understanding/
26. Adam N. Joinson. 2008. Looking at, looking up or keeping up with people?: motives and use of facebook. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, 1027–1036.
27. Cliff Lampe, Nicole Ellison, and Charles Steinfield. 2006. A face(book) in the crowd: social Searching vs. social browsing. In *Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work* (CSCW '06). ACM, New York, NY, USA, 167-170. DOI=http://dx.doi.org/10.1145/1180875.1180901
28. Lina Lee and Alfred Markey. 2014. A study of learners' perceptions of online intercultural exchange through Web 2.0 technologies. *ReCALL* 26, 3: 281–297.
29. .Hajin Lim and Susan R. Fussell. 2017. Making Sense of Foreign Language Posts in Social Media. *Proc.*

ACM Hum.-Comput. Interact. 1, CSCW, Article 69 (December 2017), 16 pages. DOI: https://doi.org/10.1145/3134704

30. Hajin Lim and Susan R. Fussell. 2017. Understanding How People Attend to and Engage with Foreign Language Posts in Multilingual Newsfeeds. In *Eleventh International AAAI Conference on Web and Social Media*.
31. Sally Maitlis and Marlys Christianson. 2014. Sensemaking in organizations: Taking stock and moving forward. *The Academy of Management Annals* 8, 1: 57–125.
32. James C. McCroskey and Thomas A. McCain. 1974. The measurement of interpersonal attraction. *Speech Monographs* 41, 3: 261–266. https://doi.org/10.1080/03637757409375845
33. Joyce S. Osland and Allan Bird. 2000. Beyond sophisticated stereotyping: Cultural sensemaking in context. *The Academy of Management Executive* 14, 1: 65–77.
34. R. Michael Paige, Helen Jorstad, Laura Siaya, Francine Klein, and Jeanete Colby. 2000. Culture Learning in Language Education: A Review of the Literature.
35. Peter Pirolli and Stuart Card. 2005. The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. In *Proceedings of international conference on intelligence analysis*, 2–4.
36. Emilee Rader, Alcides Velasquez, Kayla D. Hales, and Helen Kwok. 2012. The gap between producer intentions and consumer behavior in social media. In *Proceedings of the 17th ACM international conference on Supporting group work* (GROUP '12). ACM, New York, NY, USA, 249-252. DOI=http://dx.doi.org/10.1145/2389176.2389213
37. Dale H. Schunk. 2003. Self-efficacy for reading and writing: Influence of modeling, goal setting, and self-evaluation. *Reading &Writing Quarterly* 19, 2: 159–172
38. Tomohiro Shigenobu. 2007. Evaluation and usability of back translation for intercultural communication. *Usability and Internationalization. Global and Local User Interfaces*: 259–265.
39. Andrew D. Smock, Nicole B. Ellison, Cliff Lampe, and Donghee Y. Wohn. "Facebook as a toolkit: A uses and gratification approach to unbundling feature use." *Computers in Human Behavior* 27, no. 6 (2011): 2322-2329.
40. TextRazor. TextRazor - The Natural Language Processing API. Retrieved September 17, 2017 from https://www.textrazor.com/named_entity_recognition
41. Stefan Volk, Tine Köhler, and Markus Pudelko. 2014. Brain drain: The cognitive neuroscience of foreign language processing in multinational corporations. *Journal of International Business Studies,* 45.7, 862-885.
42. 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). ACM, New York, NY, USA, 103-106. DOI: https://doi.org/10.1145/1718918.1718938
43. Charles Wankel. 2016. Developing cross-cultural managerial skills through social media. *Journal of Organizational Change Management* 29, 1: 116–124.
44. Shaomei Wu, Jeffrey Wieland, Omid Farivar, and Julie Schiller. 2017. Automatic Alt-text: Computer-generated Image Descriptions for Blind Users on a Social Network Service. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing* (CSCW '17). ACM, New York, NY, USA, 1180-1192. DOI: https://doi.org/10.1145/2998181.2998364
45. Naomi Yamashita and Toru Ishida. 2006. Effects of machine translation on collaborative work. In *Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work*, 515–524.
46. Naomi Yamashita and Toru Ishida. 2006. Automatic prediction of misconceptions in multilingual computer-mediated communication. In *Proceedings of the 11th international conference on Intelligent user interfaces*, 62–69.
47. Kadir Yoğurtçu. 2013. The impact of self-efficacy perception on reading comprehension on academic achievement. *Procedia-Social and Behavioral Sciences* 70: 375–386
48. Dejin Zhao and Mary Beth Rosson. 2009. How and why people Twitter: the role that micro-blogging plays in informal communication at work. In *Proceedings of the ACM 2009 international conference on Supporting group work*, 243–252.
49. Bu Zhong, Marie Hardin, and Tao Sun. 2011. Less effortful thinking leads to more social networking? The associations between the use of social network sites and personality traits. *Computers in Human Behavior* 27, 3: 1265–1271.