New Generation Computing, 32(2014)163-186
Ohmsha, Ltd. and Springer

Translation Agent: A New Metaphor for Machine Translation

Chunqi SHI, Toru ISHIDA, Donghui LIN
Department of Social Informatics, Kyoto University
Yoshida-Honmachi, Sakyo-Ku, Kyoto
shi@ai.soc.i.kyoto-u.ac.jp,
{ishida,lindh}@i.kyoto-u.ac.jp

Received 5 June 2013
Revised manuscript received 6 January 2014

Abstract We present agent metaphor as a novel interactive system to promote the efficiency in machine translation mediated communication. Machine translation is increasingly used to support multilingual communication. In the traditional, transparent-channel way of using machine translation for the multilingual communication, translation errors are ignorable, due to the quality limitation of current machine translators. Those translation errors will break the communication and lead to miscommunication. We propose to shift the paradigm from the transparent-channel metaphor to the human-interpreter metaphor, which motives the interactions between the users and the machine translator. Following this paradigm shifting, the interpreter (agent) encourages the dialog participants to collaborate, as their interactivity will be helpful in reducing the number of translation errors, the noise of the channel. We examine the translation issues raised by multilingual communication, and analyze the impact of interactivity on the elimination of translation errors. We propose an implementation of the agent metaphor, which promotes interactivity between dialog participants and the machine translator. We design the architecture of our agent, analyze the interaction process, describe decision support and autonomous behavior, and provide an example of preparing repair strategy. We conduct an English-Chinese communication task experiment on tangram arrangement. The experiment shows that compared to the transparent-channel metaphor, our agent metaphor reduces human communication effort by 21.6%.

Keywords: Multilingual Communication, Machine Translation Mediated Communication, Interactivity, Translation Repair Strategy, Language Service.

§1 Introduction

Language barrier prevents people from different nations and culture to communicate with each other. To encourage business, and bring transnational cooperation, people have to overcome the language barrier. For non-foreign language learning people, they need translation support. For example, famous transnational companies, such as Facebook and Amazon will be much more successful, if their translation support is efficient. Efficient support tools continue to receive more attention.[8] Without proper support in the multilingual environment, the language barrier will make non-foreign language learning people remain on the surface, unaware of the strangeness and complexity of life beneath the waves.[30] Machine translation plays an important and promising role in the preparation of such tools. However, machine translation has limits in terms of translation quality.[32] The translation errors continue to be the barrier for machine translation mediated (MT-mediated) communication. When MT-mediated communication is used for a cooperation task, it is necessary to translate the dialog accurately, since the accurate translation of concepts is the basis of successful information transfer.[34] Even without considering the complex individual emotion-related factors, such as cultural background,[15] it is hard to deal with machine inaccurate translation in MT-mediated communication. The traditional way uses machine translators just as transparent-channels which do not facilitate users to deal with translation errors. Thus, the multilingual communication will easily break down due to the translation errors.

We propose to shift the paradigm from the transparent-channel metaphor to the human-interpreter metaphor, in view of the fact that machine translation errors cannot be ignored. The ideal of agent metaphor, originally introduced by Ishida,[9] assumes *interactivity* as a new goal of the machine translator. Interactivity is the machine initiated interaction among the communication participants; it represents the ability to take positive actions to improve grounding and to negotiate meaning.[9,10] Different from the traditional metaphor of machine translation as a transparent channel, interactivity makes it clear that translation errors are to be treated as channel noise. This noise can be suppressed through the efforts of the dialog participants. Motivated by this assumption that, the interaction between translator and users will promote the communication efficiency, the design of interactive system for multilingual communication is inspired by both the repair theory in the human-computer interaction[10,19] and the user adaptation to machine translation theory.[17] The repair theory allows agent to interact with the users to detect the miscommunication and smooth the miscommunication. We only focus on the miscommunication caused by the translation errors. There are two direct considerations following the repair of the translation errors. One consideration is when we need the repair. Do we have to initiate the repair every time, or only initiate the repair when it is needed? To answer this question, we have to decide whether the interactive system has flexible, autonomous behavior or not. The other consideration is how to implement the repair. To repair a translation error, one important way is to promote the accuracy, the other promising way is to adapt users to machine translators.

Because of the assumption of quality limitation of machine translators, we focus on the second way. The experience-based user adaptation to machine translation provides the knowledge to motivate users to conduct the repair.

We present an agent metaphor following the above interactive system design considerations. The built-in decision mechanism allows agent to motivate proper repair accordingly. Agent metaphor allows flexibly integrating existent experience or knowledge of user adaptation to machine translation. To implement the agent metaphor, we have to examine the environment of machine mediated communication. Interactivity will be influenced strongly by this translation environment. Normally, translation environments involve the translation function and the user.[4] About the former, we have to mention two characteristics of complex machine translation. One is the variable quality of machine translator output. The other is that, two messages expressing the same information can have widely different translation quality by the same machine translator. About the latter, the *activeness* of the user is ignored in the traditional transparent-channel metaphor. Activeness plays an important role in interactivity. For example, certain people get better translation results than others because they can craft expressions that suit the characteristic of that machine translator. Thus, we need careful agent designs to promote interactivity. Our strategy is to examine an experimental environment, to provide the interaction design, and to implement a prototype for this experimental environment.

We start by examining the limitation of machine translation mediated communication. We analyze the paradigm shifting, existing works on translation errors, and concrete examples. After that, by analyzing the interactions that can eliminate those error examples, we will propose an interactivity model, including the interaction process, interaction simple model and interactivity level analysis. Then, the requirements of an agent for encouraging interactivity will be studied. On one hand, the agent needs to know the translation quality. On the other hand, the agent needs to adapt users to the machine translator. Furthermore, we will describe the design of the agent metaphor, including its architecture and functions. In order to evaluate our prototype, we conduct a wizard of OZ experiment on the multilingual tangram arrangement task. Finally, we summarize what has been learned before discussing the limitations and implications of the current design.

§2 Limitation of Current Machine Translation Mediated Communication

We want to explain the limitation of current MT-mediated communication from the necessary of turning the direction from accuracy promotion to interactivity motivation, the categories of existing work on the repairs of translation errors, and the concrete examples translation errors, which will motivate the analysis of interactivity.

2.1 From Accuracy Promotion to Interactivity Motivation

Machine translation mediated communication is the computer mediated communication which applies machine translation to facilitate multilingual communication. Accuracy is the goal in the traditional applying of machine translation. However, inaccurate translations are ignorable and break down the multilingual communication. This goal of accuracy does not count in the users, without considering any linguistic and social nature of the users. Thus, MT-mediated communication is viewed as a transparent channel for information transfer, and translation errors will lead to miscommunications.

After the birth of the human-interpreter metaphor of MT-mediated communication, the importance of interactivity is stressed.[9] In computer mediated communication, interactivity is to understand the influence of responding to a message by studying the relationship between messages in human to human communication and then human to machine communication.[25] Its meaning is refurbished after viewing MT-mediated communication as the conversations between machine translator and users of different languages. Interactivity is defined as the ability of machine translators to take actions to improve the grounding and to negotiate meanings in the multilingual communication.[9] It becomes the new goal in MT-mediated communication and the linguistic and social nature of the users should be regarded in the human-interpreter metaphor. For example, the level of user's foreign language skill will affect this multilingual communication. Then, we face the problem that how we should implement the human-interpreter metaphor to motivate interactivity, so that the number of communication breakdowns caused by translation errors will be reduced.

The emergence of repair theory brings promise to the implementation of human-interpreter metaphor to motivate interactivity. Repair in MT-mediated communication is the user actions to raise the quality of the machine translation output, so as to renovate miscommunication.[13] An increasing number of repairs have been studied as the availability of machine translation increases. Especially, the birth of language services greatly promotes the usability of general machine translations, such as language services from the Language Grid, a service-oriented platform.[11] Thus, to turn from the accuracy promotion to the interactivity motivation, we want to implement the human-interpreter metaphor as a translation agent, which is designed to interact with users to motivate the repairs of translation errors with support from various language services.

2.2 Existing Work on Translation Errors

We examined existing works on translation errors from the user perspective. Translation environment involves both translation function and user. From the perspective of translation function, the analysis of machine translation errors is very important for the development of machine translation.[23] Here, from the perspective of users, the analysis of translation errors is different in whether user can use it to manually correct translation or not. In MT-mediated communication, translation errors lead to miscommunication. Analyzing miscommunication at the phrase, sentence, and dialog level is popular in machine-mediated communication research.[14,34] These three observations of machine translation errors are picked up according to these levels: *phrase-level*, *sentence-level*, and *dialog-level*.

Table 1 Existing works on three levels and their corresponding mistranslation problems.

Level	Existing Work	Translation Errors
Phrase level	Extract and highlight inaccurate words,[19] picture icons as precise translation of basic concepts.[28]	Inadequate
Sentence level	Examine back-translation for sentence level accuracy check,[18] Round-trip monolingual collaborative translation of sentence.[6, 20]	Disfluent and inadequate
Dialog level	Examine asymmetries in machine translations,[35] Predict misconception due to unrecognized translation errors.[34]	Inconsistent

Table 1 shows several existing works on examining mistranslation problems, providing suggestions and strategies for improving accuracy at each level. For example, in phrase level, highlighting inaccurate words will facilitate user modification. In sentence level, back-translation will provide certain information of translation result. In dialog level, prediction of potential translation inconsistency prevents user using an improper shortened reference of the previous concept. We summarize the mistranslation found in existing works. It not only shows that mistranslation often happens and can lead to communication breaks, but also there are existing works on user perspective dealing with variable situations of translation errors.

2.3 Examples of Translation Errors

To show the translation errors leading to miscommunication, we will give one concrete example for each of the phrase, sentence, and dialog level. These translation errors are picked out from English-Chinese machine translation mediated dialogs, in which the English instructs the Chinese to arrange tangram figures.[*1] Using these examples, we show the communication breaks occasioned by different translation errors. In first example, due to phrase concept mistranslation, the Chinese receiver cannot understand. The word "square" in the geometry domain was translated into "plaza" of another domain, because the word "square" is polysemous. The machine translator just provides the everyday meaning of the word, but its true meaning depends on the domain of communication task. In the second example, the mistranslated sentence is an imperative sentence that requests the receiver to conduct an act ("put something someplace"). This sender describes actions in imperative sentences, such as requests and commands. The last example (see Fig. 1) is the mistranslation of inconsistent phrases. The abbreviated reference ("the light one") is not translated accurately, and it is unnatural to stick to exactly the same expression globally. Such inconsistency easily leads to translation errors.

From above, machine translators cannot guarantee the translation accuracy; communication will break due to those translation errors. There are many

[*1] http://en.wikipedia.org/wiki/Tangram

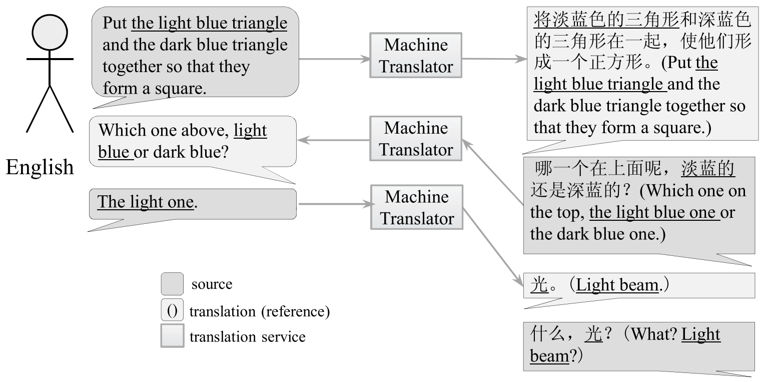


Fig. 1 In English-Chinese tangram arrangement communication, the Chinese receives an inconsistent translated phrase and the communication breaks.[*2] (In this figure or following figures, the parentheses indicate references of Chinese-to-English translation.)

existing works on various translation errors in phrase, sentence and dialog levels. From the examples on each level, the translation errors will not be interpreted by the users, and thus break the multilingual communication.

§3 Interactivity Modeling

We will show how interactivity could be helpful to reduce the number of translation errors. After that, we will design a unified interaction process. Based on the intent analysis of interactions, we will propose a simple interaction model and level up its interactivity.

3.1 Interactions for Repairing Translation Errors

Interactions between machine translator and users could help the users repair the translation errors. Communication "repair" is the word referring to the human effort to fix the communication breakdown. In MT-mediated communication, "repair" has been used to refer the human effort to fix translation errors.[10, 19] Adapting users to machine translation has been proposed in existing works (see Table 1). Those were specific repairs to particular situations. We want to study a general interaction process for repairing MT-mediated communication. To solve the mentioned translation errors in the last section, we show interactions to repair those translation errors. When translation errors could be eliminated through collaborating with the users, the goal of the machine translator becomes to encourage interactions. We studied the process of interactions, which eliminate the translation errors.

We introduce three interactions motivating users to repair the translation error examples. In the first repair (see Fig. 2), the "square" is main direct object and polysemous. Based on the dictionary, the multiple target concepts were sent

[*2] The square is with 45 degree angles to the horizontal line, so that one triangle is above the other one.

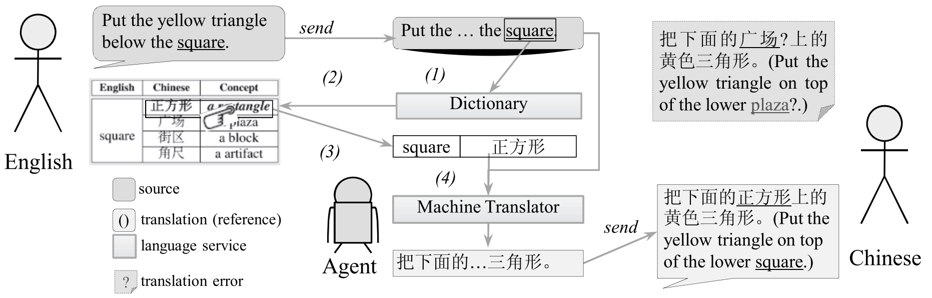


Fig. 2 Interaction to handle inadequately translated phrase: *(1)* check the feature that the word, "square," has one-to-many dictionary results. *(2)* suggest the sender select the correct concept. *(3)* the sender chooses the target concept. *(4)* translate by the dictionary translator composite machine translator.

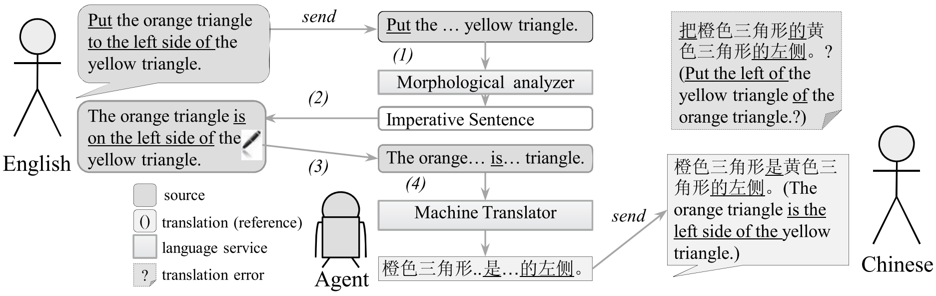


Fig. 3 Interaction to handle mistranslated sentence. *(1)* check the feature that it is an imperative sentence starting with a verb. *(2)* suggest the sender rewrite the sentence into declarative version. *(3)* rewrite the sentence. *(4)* translate by the machine translator.

to the sender, who selected one target concept. The target phrase was integrated into the translation. This translation error can also be solved through existing works on domain adaptations and disambiguation.[33] The difference is that this interaction is an online model and user-oriented, while the domain adaptations are offline models and machine-oriented works. In the second repair (see Fig. 3), based on the morphological analyzer, the sentence was detected as an imperative sentence. Thus, the sender was suggested to rewrite it into declarative version, because machine translators often fail to translate imperative sentences as well as declarative sentences. The last repair (see Fig. 4) suggested a phrase memory to keep and retrieve the similar phrases. An automatic suggestion for replacement can solve the inconsistency of varied reference problem.

3.2 Interaction Process

Once unacceptable translation is detected, we propose an interaction process to motivate users to repair the translation errors. The concrete interactions for motivating users to conduct repair will be called as *repair strategy*. Given that there are multiple repair strategies, the agent has to decide the cause of

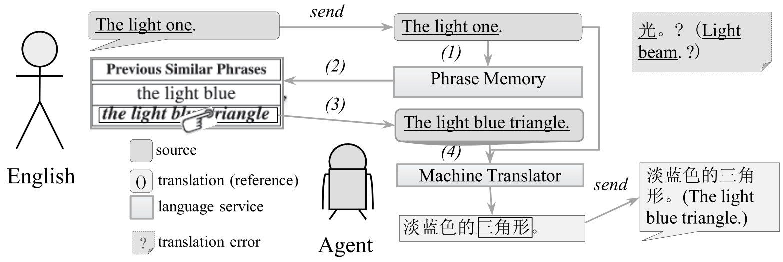


Fig. 4 Interaction to handle inconsistently translated phrase. *(1)* check the feature of similar phrases existing in previous dialog. *(2)* suggest selection of appropriate previous phrase. *(3)* choose a replacement of the previous phrase. *(4)* translate by the machine translator.

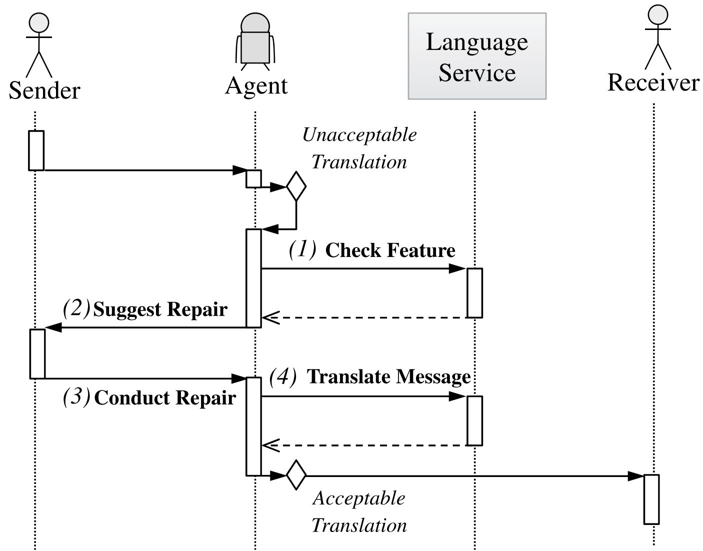


Fig. 5 Four steps of the interaction process for one repair strategy: *(1)* Check Feature, *(2)* Suggest Repair, *(3)* Conduct Repair (Sender), and *(4)* Translate Message.

the failure, send the appropriate repair suggestion to the sender. Other types of repair strategies, such as selecting phrases based on the prediction of available information,[4)] rephrasing based on back-translation results and sentence rewriting,[19)] can also be used. Thus, a unanimous interaction process is necessary to integrate different repair strategies. The proposed interaction process consists of four steps (see Fig. 5) :

1. Agent's move to check the *feature* of current message.
2. Agent's move to *suggest* repair tips to the sender.
3. Sender's move to conduce the repair.
4. Agent's move to translate the repaired message, and output an improved

translation result.

Obviously, if the agent can initiate proper interactions, most translation errors can be eliminated. Still, we have to mention that the sensibility of users does not necessarily lead to the elimination of translation errors, because of the unpredictability of the machine translation function, and the uncertainty of the human repair action. Thus, the interactivity between the agent and dialog participants must be carefully designed to motivate users by making their actions easy, even for monolingual neophytes.

3.3 Simple Three-Interaction Model

Accordingly, we propose simple three-interaction model (see Fig. 6). From the above interactions, the interaction message varies among different repairs, but the *intent* of these interactions is clear. We describe three typical categories of interactions based on the analysis of their intent. The typical categories include:

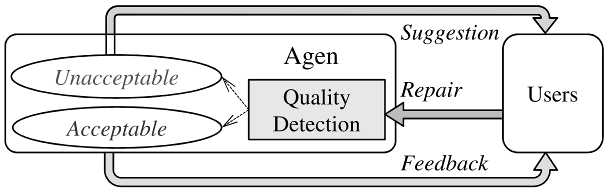


Fig. 6 Three-interaction modeling of agent metaphor for machine translation mediated communication.

1. *Suggestion*: when translation is found unacceptable, the repair suggestion will be sent to the participants. As mentioned previously, more than one repair can be available. To initiate a suggestion, the repair strategy has to be decided and the tips should be prepared.
2. *Repair*: the users conduct the repair to resolve the miscommunication. It can be an adaptation to machine translation,[13)] such as pre(post)-editing,[24)] or other ways to improve translation quality, such as paraphrase word,[3)] split sentence, or even ways to improve understanding, such as the redundant[27)] or supplementary information (e.g. pictogram[31)]). It should be noticed that the users can conduct the repair without any suggestion. It can be an agent-initiated repair or a user-initiated repair.
3. *Feedback*: after the quality detection, the quality can be either unacceptable or acceptable, and it will be fed back to users. Back-translation is often used for the quality evaluation.[18)] Other quality evaluations can also be considered.[29)]

From the modeling, it is obvious that built-in translation quality detection is important for understanding whether translation quality is acceptable or not. Without quality detection, it is difficult for the agent to notify users the translation quality, not to mention motivating them repair the errors. Besides, it allows the reiteration of these interactions. In reiteration, the repair suggestion implies the quality.

3.4 Interactivity Levels

In the above three-interaction model, we find category-typical *intents* by separately analyzing each interaction in one direction in the process of repairing translation errors. As a whole process, we also analyze the *initiation* of variable repairing processes and we propose interactivity levels based on this analysis. The ability of agent in motivating users to conduct a repair is tightly related to the initiation. We clarify three typical interactivity levels (see Fig. 7), which is inspired by the taxonomy of interactivity from the work of Schwier:[26)]

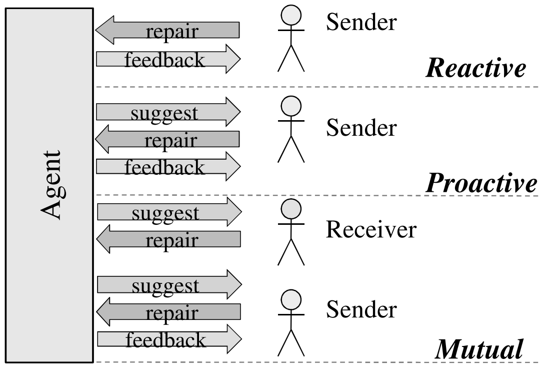


Fig. 7 Interactivity Levels: Reactive (participant-initiated repair), Proactive (agent-initiated repair), and Mutual (agent-initiated collaborative repair)

1. Reactive: it is the user-initiated repair. The agent gives feedback on the quality after repair. It depends on user's experience in repairing translation errors. Because the agent does not have to interpret and initiate the repair suggestion, we call it a reactive level. Agent could feedback the quality of the user-initiated repairs.
2. Proactive: it is the agent-initiated repair. The users listen to and follow the suggestion to conduct the repair. To provide a suggestion, agent has to interpret the translation error, decide the repair strategy and give tips. It is called proactive level, because agent guides the repair process.
3. Mutual: it is an agent-initiated collaborative repair. Agent will initiate interaction to both sender and receivers. The agent will provide suggestion and activate collaborative repair. It represents agent's ability to coordinate both the receiver and sender together.

An example in the reactive level is to repair the mistranslation of the polysemous word "square." After knowing that the quality is unacceptable, the sender initiated the repair rephrasing the word using a more specific word "foursquare." Obviously, such rephrase action indeed requires the knowledge or experiences from the sender. The sender has to decide the proper repair upon his/her own knowledge. An example in the proactive level is to repair the mistranslation of the imperative message "*Put the orange triangle to the left side of the blue triangle*" (see Fig. 3). Agent detected the feature "an imperative sentence," and decided the repair strategy rewriting it into declarative version. The sender received this suggestion and conducted the rewriting. An example in the mutual level is to repair the same mistranslation of the imperative message. The repair is divided into two subtasks that 1) the agent asks the receiver to concrete the translation problem "It is not clear which is on the left" and 2) the agent asks the sender to rewrite message using the receiver's information "The orange triangle is on the left side of the blue triangle."

The interactivity level will be helpful in understanding existing related works. For example, the self-initiate repair,[18] which well falls into the reactive level of our proposal, shows its weakness in the face of the changes of translation function. The repair support agent[19] proposed a back-translation mechanism to deal with the anticipatory inadequacy. It belongs to the proactive, but with limited repair methods available. Another is collaborative translation,[21] which has protocol to coordinate both users. Still it expects changes[7] and better easiness to motivate users. It provides a coarse-grain description of agent's ability.

Moreover, the interactivity level implies simple user modeling for human-interpreter agent. In MT-mediated communication, the users are often described as experienced/novices user, and bilingual/monolinguals. The reactive, proactive, mutual levels have ability to deal with bilingual/monolingual experienced users, bilingual/monolingual novice users, monolingual novice users. Considering such user modeling and the built-in translation quality evaluation, a proper translation agent can be designed with concrete agent-participant interactions to elicit the repair suggestion with machine translation techniques.[22]

§4 Agent Metaphor for Interactivity

We propose an agent metaphor for the interactivity. According to the interactivity model, the proactive interactivity needs agent's initiation to repair translation error. We have proposed the interactions process of applying repair strategies. Thus, we will design the agent to initiate the repair using the interaction process. We discuss why the agent metaphor is needed to establish the proactive interactivity. Basically, to apply the agent metaphor, we have to define the agent sophistication and the role of the agent.[12] In this study, we want to clarify them through *flexible autonomous behavior* and *decision support function* of our agent metaphor.

Flexible Autonomous Behavior: An agent enables flexible autonomous translation and repair, which is much more efficient for multilingual communication. Because online translation and interactions need variable actions, such as dictionary retrieval or morphological analysis, a proactive agent has the ability to avoid unnecessary operations. On the contrary, for example, process protocol based collaborative translation[7,20] is potentially inefficient to go through the complete preset process flow.

Decision Support Functionality: Decision support enables dealing with complex environment of MT-mediated communication, such as translation error candidates, repair suggestions or extra translation improvement actions. A simple premise of such decision should be drawn from current translation environment. Through further design enhancement, the agent metaphor will gather not only translation accuracy, but also additional quality estimates or information from the users. Thus, the agent metaphor has to sense the environment like the quality of current translations, build common awareness among users about current translation, and pass proper repair suggestions to users.

4.1 Architecture Design

We will focus an agent metaphor for the proactive interactivity, because this level covers all the interaction categories: suggestion, repair, and feedback. Once we have the agent metaphor for the proactive interactivity, it will be adapted to other two levels. Our translation agent is designed around three agent phases: observation, decision, and action (see Fig. 8). Observation phase will perceive the translation quality, which is needed for the decision phase. Decision phase will make a plan of actions. The action phase will follow the plan and fulfill the actions.

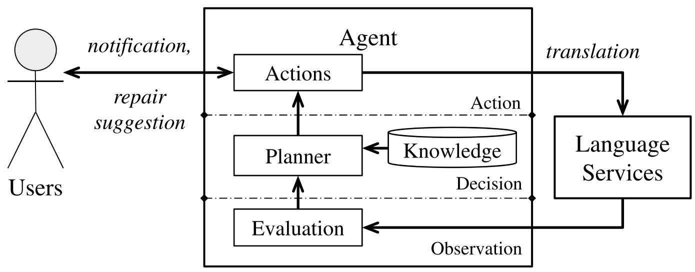


Fig. 8 Architecture Design of Translation Agent

[1] Observation phase

The translation quality will be discerned by an *evaluation* module and it will be passed to the next phase. The translation quality contains not only the accuracy but also quality features. The accuracy will be used for decision of translation acceptable or not. While the quality feature will be used for decision of taking which repair strategy. For accuracy, popular evaluation methods such as BLEU, METEOR, compare the lexical similarity between the translation result and a standard reference to calculate an evaluation score. Previous studies use back-translation to predict potential translation errors.[7, 18, 20] In this paper, back-translation and the BLEU method (maximum 3-gram, smoothed) are used as a simple way to estimate accuracy. About quality features, many quality estimation approaches can be considered.[29] In this paper, we develop own quality features based on the development of repair strategy, which will be mentioned in next section.

[2] Decision phase

A real time *planner* is necessary for online communication. A knowledge base is needed to keep experience and/or policy. The planner is critical to establishing autonomous behavior and decision support, by describing the experience, policy, or decision in rules. The activities of the users will provide uncertain results. This is because the users might have limited ability to generate correct repair actions or the machine translation quality of each message is unpredictable. Accordingly, the planner should provide online planning and decision support to counter this uncertainty. The knowledge base will save and allow access to experience and policy.

[3] Action phase

Three types of actions are needed. First, to help the users get an idea of current quality, a *notification* action is needed. Second, the detection of an unacceptable translation triggers the *repair suggestion* action. The repair suggestion is the key to interactivity. Last, *translation* actions are needed to implement the different repair strategies. Both notification and repair suggestion demand that the agent and dialog participants talk. We use a simple meta-dialog for this purpose. It includes five meta-dialog acts: INFORM, SUGGEST, REPEAT, ACCEPT, and CANCEL. This means that the agent can inform the sender and receiver status of detected translation quality, pass repair tips to the sender, repeat the process until accept the repaired message or terminate the process if improvement proves to be impossible. The meta-dialog rule, <IF, THEN, PRIORITY>, is enough to trigger one meta-dialog and properly sequence the several triggered meta-dialogs. Based on the five meta-dialog acts, the agent is able to finish the actions of notification and repair suggestion.

For the actions of notification and repair suggestion, the demand is that the agent and dialog participants talk. We use a simple meta-dialog for this purpose. For the translation actions, the repair strategies in the observations of the last section require the dictionary service result, and the dictionary translator composition service (see Fig. 2). These services will be provided through the Language Grid, a service-oriented platform.[11] Through Language Grid, several categories of atomic language services are available, including dictionary, parallel text, morphological analyzer, dependency parser, machine translators, etc. Meanwhile, several composite services are available, including dictionary-translator composite translation, multi-hop machine translation, and back-translation. Language Grid also provides a client that supports the invocation of both atomic and composite language services. People can develop their own version of services based on this client using Java programs. Language Grid platform support allows translation actions to be realized and invoked flexibly.

4.2 Autonomous Behavior and Decision Support

Sharing the status of translation quality between participants, and helping participants adapt to machine translation, are two goals of interactivity. Each communication dialog consists of many rounds of message transfer from one participant to the other. Through this transfer, the agent triggers interactivity. There are two message transfer states: *Acceptable* accuracy, and *Unacceptable* accuracy. If the former, after the message is translated into the other languages, and the accuracy is accepted, the translated message is sent to the receiver. If the latter, the agent will notify the participants and pass repair tips to the sender who then repairs the message. The message will be sent to the agent again, and the message transfer process repeats.

Second, two interactivity goals should be met. Satisfying the first goal, sharing the status of translation quality between participants, is obvious. In the above Unacceptable accuracy situation, an informational meta-dialog will be triggered and a notification meta-dialog message will be sent to the sender. A decision on whether it is acceptable or unacceptable is needed. For the second goal, helping participants adapt to machine translation, achieving the goal is essential. Based on the previous case study of interactivity, we learned three points. The first point is that there is more than one repair strategy. This means that the agent has to decide which strategy should be taken. The second point is that repair is a four-step process {*feature*, *suggest*, *conduct*, *translate*}. The third point is that the effect of any repair action is uncertain. The decision, deciding which repair strategy is to be selected under uncertainty, is especially important.

The agent has to decide whether to pass the message to the receiver,

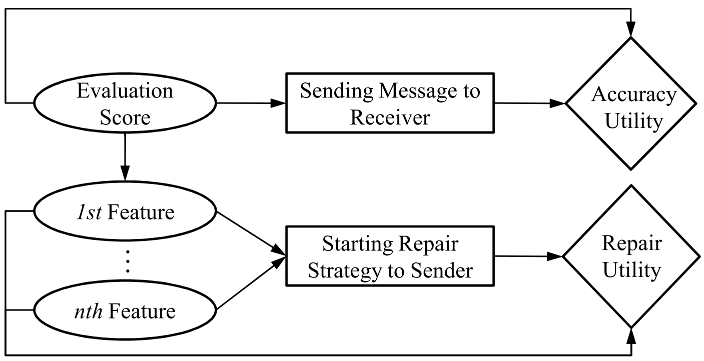


Fig. 9 Influence diagram of agent's decision dependency. One decision is for whether to send message to the receiver, and the other is which repair strategy to pass to sender. (Here, the oval node represents chance node (random variable), and its in-arcs represent probabilistic dependence. The square node represents decision node of actions, and its in-arcs represent information available when the decision is made. The diamond node represents utility node, and its in-arcs represent utility dependence.)

and if not, which repair strategy is to be taken. Through the influence diagram (see Fig. 9) of agent, we describe the decision dependency present in our agent metaphor. When the message is received, translated, and evaluated, the evaluation score is calculated via back-translation. The evaluation score determines whether the message is passed on or a repair strategy is needed. About the second decision, which repair strategy to adopt, the features of multiple repair strategies are checked and one is selected. Based on the influence diagram and message transfer situation, the utility decision model can be used to meet this two decision requirement.[2)] This autonomous behavior and decision support allow the agent to issue the appropriate repair strategy even under uncertainty.

4.3 Repair Strategy Example

An example of issuing the repair strategy "split," is explained. Here, we picked one rule from the AECMA Simplified English Standard,[1)] which is for technical manual preparation, and tried using it as the basis of a repair strategy, because simplified writing is effective in enhancing machine translation quality according to Pym's study.[24)]

Simplified Writing Rule: use short sentences. Restrict sentence length to no more than 20 words (procedural sentences) or 25 words (descriptive sentences). Inspired by this rule, we developed the repair strategy "split."

Repair Strategy Split: when an unacceptable translation is detected, if the message is a long and complex sentence, the repair tip is to split the source sentence into two sentences.

Feature of Split Strategy: the literal length of the sentence is not directly used here. Instead, we choose the syntax-tree-width of its non-leaf syntax tree (see Fig. 10). For example, the English message from the tangram arrange task, "Please place the blue triangle upside down into the middle of two given triangles." is parsed into a constituency structure tree. The non-leaf part nodes form

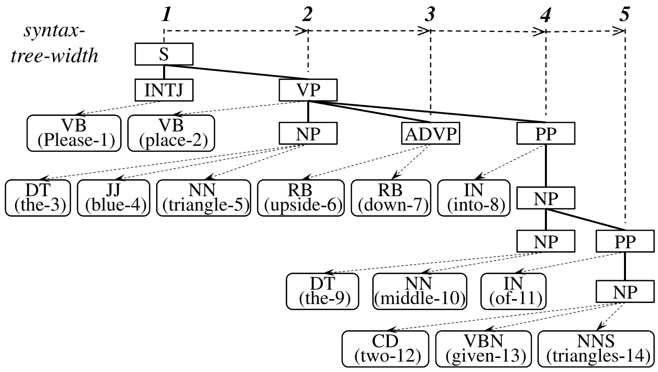


Fig. 10 The *syntax-tree-width* feature of the repair strategy *split*: a width of non-leaf part of its constituency structure tree (abbreviations are from Stanford part-of-speech tagger[*3]).

[*3] http://nlp.standford.edu/software/tagger.shtml

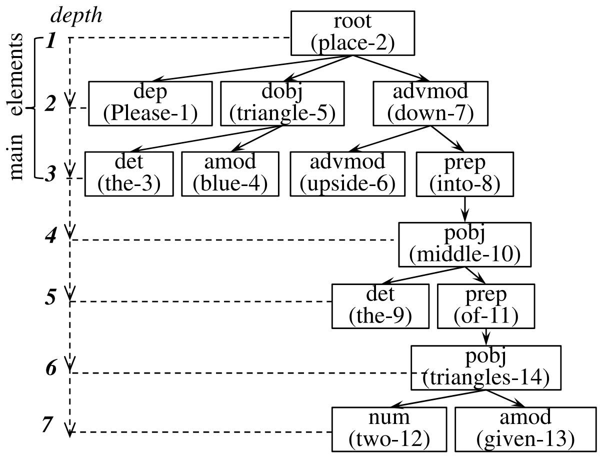


Fig. 11 The *tips* for the repair strategy *split*: the core of the message, defined as the main elements of low depth (less than 4) in the dependency structure tree (abbreviations are from Stanford typed dependencies[*4]).

a non-leaf syntax, and its width is 5. Compared to the literal message length, this syntax-tree-width better represents the complexity of sentence structure.

Repair Suggestion: the tips are provided to help the sender undertake the repair. In this repair strategy, the core of the message, which is the main elements of the sentence with low depth (less than 4) in the dependency structure tree, is picked out for the sender (see Fig. 11). This meta-dialog shows that, if the repair strategy is "split," then the suggestion and repair tips are passed to the sender (see Fig. 12). The priority value is 0.5. It means that this will be the first message shown to the sender, if there is no higher priority meta-dialog defined for the IF premise. Both the constituency parse tree and dependency parse tree are from Stanford Parser,[16)] which is an open source Java implementation of natural language parsers. It provides a consistent interface for dependent parsers for English, Chinese, Arab, and German.

Here we describe the process of preparing our split repair strategy. According to our observation of the English-Chinese tangram arrangement sessions, we found instances in which this repair strategy was needed (see Fig. 12). Obviously, the translated message is initially evaluated as unacceptable. We use the back-translation, the BLEU score, and the threshold for a simple decision. The usage of back-translation has been discussed a lot.[7, 18, 20)] The interaction process of split strategy is given and its result is shown (see Fig. 12). The agent checks the feature of split strategy, prepares the suggestion tips, and feeds back the split suggestion. After the sender splits the message following the tips, the translation is evaluated again and it becomes accepted.

*4 http://nlp.standford.edu/downloads/dependencies_manual.pdf

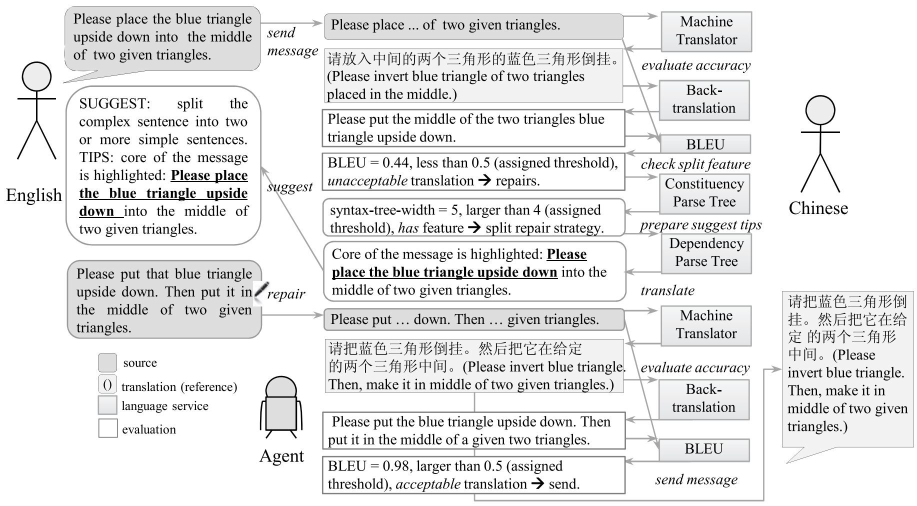


Fig. 12 Example of agent's split strategy. Agent will follow the decision of sending message to receiver or starting repair strategy. Back-translation and BLEU evaluate translation quality (*accuracy*). The process steps of the repair strategy (see Fig. 5) are given: check split feature, *the syntax-tree-width* (see Fig. 10); suggest split with tips, *the core of the message* (see Fig. 11); human repair; and translate the repaired message. Then, the evaluation shows that it becomes accepted.

§5 Evaluation

In order to evaluate the impact of the agent metaphor, we conducted a controlled experiment, tangram arrangement, to compare the proposed agent metaphor approach with the machine translator mediated transparent channel approach. The reason of choosing tangram arrangement is that it transfers simple information that we can easily calculate the number of messages. Since the translation agent prevents the transfer of translation errors and reduces the number of user messages needed to address the translation errors, we can calculate the average number of user messages in transferring each piece of tangram as the communication costs.

5.1 Experiment Preparation

An English-Chinese tangram arrangement communication task was conducted: an English user instructs a Chinese user how to arrange a tangram (see Fig. 13). When the tangram is complex, this task is generally difficult to finish through text based messages, even for two native speakers. We set two limitations to make this task easier to finish. *Only use convex figures*:[*5] there are only 13 convex figures. It is much easier to construct a convex figure. *Share initial state of tangram pieces*: both participants start with the same piece arrangement. With these two limitations, tangram arrangement focuses on communication.

[*5] http://en.wikipedia.org/wiki/Tangram

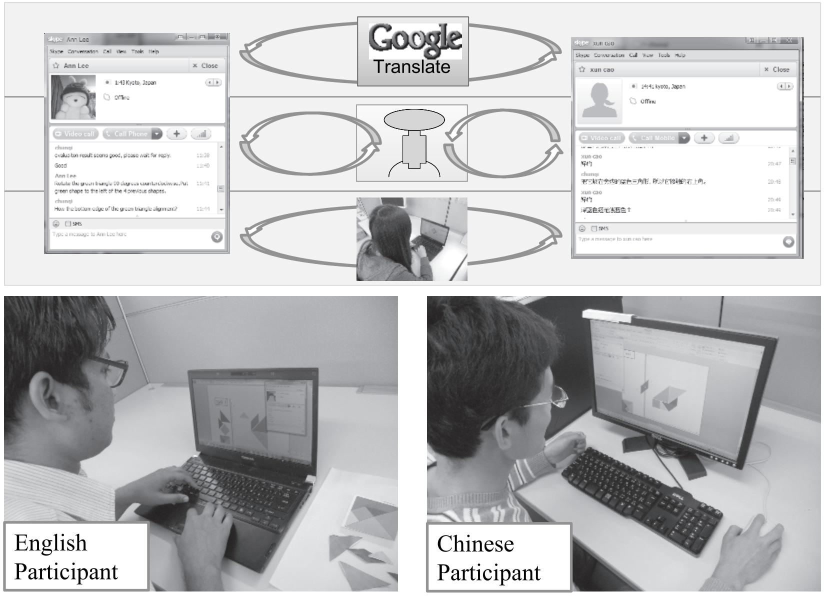


Fig. 13 Experziment of English-Chinese (E-and-C) tangram arrangement through machine translator (MT), agent prototype using the wizard of OZ (Agent), and human bilingual (Human). There are two groups of participants, E_1-and-C_1 and E_2-and-C_2 for this tangram arrangement experiment.

We examine the elimination of translation errors raising the efficiency of communication. Higher efficiency means that the information is transferred with fewer messages. According to conversation analysis,[5)] the *turn* is the basic unit interaction in the communication. The tangram arrangement task can be divided into seven subtasks; there are seven pieces to be arranged. For each arrangement, the information transferred per turn unit, includes piece, rotation type, and position. The *number of human messages per turn unit* is defined as the number of messages sent by the human participants during one turn unit of the multilingual communication. It reflects the participants' effort to transfer the task information. For better data collection, after one message is sent, the participants were asked to wait for feedback before issuing the next message.

Normally, a turn unit consists of two messages: one information message from the sender and one feedback message from the receiver. Here, to transfer the square's position information, 4 messages are needed (the number of human messages is 4) because the translation error misleads the message receiver, and the receiver has a query. It should be noted that, in the agent metaphor, the repaired message from the sender is counted, for example, the number of human

ngc32204 : 2014/4/1(9:27)

messages in the turn unit is 2 (two messages from the English sender) in the split strategy example (see Fig. 12).

For each tangram, we conducted the task using a single machine translator, a translation agent prototype, and bilingual translators. We randomly selected 5 tangram figures from the 13 convex figures. 2 English and 2 Chinese participants, and 1 English-Chinese bilingual joined this experiment. Two groups are formed with 1 English and 1 Chinese in each group. In this experiment, the agent prototype knew three repair strategies; the *split* strategy of the last section, and the two repair strategies of Fig. 2 and Fig. 3: *phrase* and *rewrite*.

5.2 Result and Analysis

Each group (1 English and 1 Chinese) was asked to finish 5 figures. The number of human messages and the average number of human messages in each turn were collected (see Table 2). The average number of human messages in each turn in human-mediated communication is 2.2. This shows that human-mediated communication is pretty efficient. The average number of human messages in each turn in machine translator mediated communication was 3.7. This shows that using machine translation almost doubles the participants' effort. Our prototype agent held the average number of human messages in each turn to 2.9, a **21.6%** improvement in communication efficiency.

Next, the total number of repair strategies in the English-Chinese dialogs was determined (see Table 3). First, the two different message senders had different repair strategies. Sender E_2's messages triggered more repair suggestions.

Table 2 The total number of human messages and the average number of human messages in each turn unit of the 5 English-Chinese communication tasks.

NO.	**Medium**	**Total Number of Human Messages**		**Average Number of Human Messages / Turn**
		E_1**-and-**C_1	E_2**-and-**C_2	
1	MT	28	26	3.9
	Agent	21	19	2.9
	Human	18	16	2.4
2	MT	28	28	4.0
	Agent	23	24	3.4
	Human	16	16	2.3
3	MT	24	24	3.4
	Agent	20	19	2.8
	Human	14	14	2.3
4	MT	24	24	3.4
	Agent	18	19	2.6
	Human	16	14	2.1
5	MT	26	24	3.6
	Agent	19	18	2.7
	Human	14	14	2.0
Avg.	MT	26.0	25.2	3.7
	Agent	20.2	19.8	2.9
	Human	15.6	14.8	2.2

Table 3 The Total Times of the Repair Strategies

Sender	Total Times of the Repair Strategies		
	Phrase	**Rewrite**	**Split**
E_1	6	21	9
E_2	5	17	10

The phrase and split strategies were used to almost the same extent. Second, different repair strategies took different amounts of time to complete. Here, the phrase strategy and split strategy were not activated as frequently as rewrite. This might be because there were few polysemous words, and the sentence structures were not too complex. We note that the senders used many imperative messages in the first instance.

§6 Discussion and Conclusion

Before concluding, we want to consider the scalability and limitation of the proposed agent metaphor. Only one controllable experiment has been studied here, but the agent metaphor is applicable to other multilingual communication, for example, school orientation to new foreigner students. The agent metaphor has provided the framework to integrate existing work on manually repair translation errors, including the architecture, mainly components, interaction process, etc. Of course, a simple implementation with several repair strategies, such as phrase strategy and rewrite strategy, will help communication users. For more general agent for any multilingual communications, current research has many limits on such as larger amount of repair strategy collection, more flexible experience analysis and training on the planner, and more intensive user modeling on monolingual or foreign language learning users. But for higher promotion of communication efficiency, mostly, customization is necessary for different multilingual communication task and different users. For example, learning to decide the best repair strategy using translation errors history of similar tasks. For certain multilingual communication task about road directions, certain repair strategy such as pictograph[31)] or arrow diagram can be developed. For another example, adapting the repair suggestions and interactivity levels to different users. For certain foreign language learners, when they have repaired several translation errors according to the suggestion from agent, they can develop their own experience on repairing, and own decision on available repair strategies. For such special situations, a reactive level interactivity can be a better choice than the proactive level interactivity. Thus, we provided the framework of agent metaphor, which allows customization on different multilingual communication application to gain better communication efficiency.

In summary, we proposed the agent metaphor based on the analysis of interactivity model, which represents a paradigm shift from transparent metaphor to human interpreter metaphor. We analyze the limitation of machine translation mediated communication. We examined the phrase, sentence, and dialog level translation errors, and showed that interactions could support the users in eliminating translation errors efficiently. Then, we proposed an interaction process, three-interaction model and interactivity level. After that, we proposed the agent architecture and the lessons to build up a repair strategy. To realize the autonomous agent mechanism, the process of repair suggestion was analyzed, the situations of message transfer were described, and decision dependency was analyzed for autonomous behavior and decision support. Finally, we described our English-Chinese communication experiments on a tangram arrangement task. The results showed that the wizard-of-OZ prototype of our agent metaphor improved communication efficiency in the face of translation errors. The agent does save users' effort to solve translation errors to promote the communication efficiency.

Acknowledgements

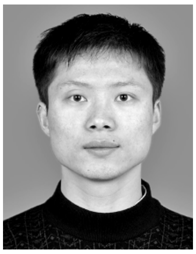
This research was partially supported by Service Science, Solutions and Foundation Integrated Research Program from JST RISTEX, and a Grant-in-Aid for Scientific Research (S) (24220002) from Japan Society for the Promotion of Science. We are very grateful to Ann LEE, Xun CAO, Amit PARIYAR, Mairidan WUSHOUER, Xin ZHOU for their helps in the experiments.

References

1) AECMA, *A Guide for the Preparation of Aircraft Maintenance Documentation in the Aerospace Maintenance Language*, AECMA Simplified English, Brussels, 1995.
2) Bohnenberger, T. and Jameson A., "When policies are better than plans: decision-theoretic planning of recommendation sequences," in *Proc. of IUI2001*, ACM, pp. 21–24, 2001.
3) Callison-Burch C., Koehn, P and Osborne M., "Improved statistical machine translation using paraphrases," in *Proc. of the main conference on Human Language Technology Conference of the North American Chapter of the ACL, HLT-NAACL'06*, pp. 17–24, 2006.
4) Carl, M., Way, A. and Schäler, R., "Toward a hybrid integrated translation environment," in *Machine Translation: From Research to Real Users, LNCS 2499*, Springer Berlin / Heidelberg, pp. 11–20 2002.
5) Goodwin, C. and Heritage, J., "Conversation analysis," *Annual Review of Anthropology*, *19*, pp. 283–307, 1990.
6) Hu, C., "Collaborative translation by monolingual users," in *Proc. of the EA2009*, ACM, pp. 3105–3108, 2009.
7) Hu, C., Bederson, B. B., Resnik, P. and Kronrod, Y., "Monotrans2: a new human computation system to support monolingual translation," in *Proc. of CHI2011*, ACM, pp. 1133–1136, 2011.
8) Inaba, R., "Usability of multilingual communication tools," in *Usability and Internationalization*, *Global and Local User Interfaces, LNCS 4560*, Springer Berlin / Heidelberg, pp. 91–97 2007.
9) Ishida, T., "Communicating culture," *IEEE Intelligent Systems*, *21*, pp. 62–63, 2006.
10) Ishida, T., "Intercultural collaboration using machine translation," *IEEE Internet Computing*, pp. 26–28, 2010.

11) Ishida, T., *The Language Grid: Service-Oriented Collective Intelligence for Language Resource Interoperability*, Springer, 2011.
12) Jennings, N. R. and Wooldridge, M., "chapter Applications of intelligent agents," *Agent technology*, Springer-Verlag, New York, Inc., pp. 3–28, 1998.
13) Nomura, S., Ogura, K., Hayashi, Y. and Ishida, T. "User adaptation in mt-mediated communication," in *The First International Joint Conference on Natural Language Processing*, pp. 596–601, 2004.
14) Kiesler, S., Zubrow, D., Moses, A. M. and Geller, V., "Affect in computer mediated communication: an experiment in synchronous terminal-to-terminal discussion," *Hum.-Comput. Interact.*, *1*, *1*, pp. 77–104, March 1985.
15) Kim, K-J., "Cross-cultural comparisons of online collaboration," *Journal of Computer-Mediated Communication*, *8*, 2002.
16) Klein, D. and Manning, C. D., "Fast exact inference with a factored model for natural language parsing," in *Advances in Neural Information Processing Systems (NIPS), 15*, MIT Press, pp. 3–10, 2003.
17) Lehmann, S., Gottesman, B., Grabowski, R,, Kudo, M., Lo, S., Siegel, M. and Fouvry, F., "Applying cnl authoring support to improve machine translation of forum data," in *Controlled Natural Language*, *LNCS 7427*, pp. 1–10, 2012.
18) Miyabe, M. and Yoshino, T., "Accuracy evaluation of sentences translated to intermediate language in back translation," in *Proc. of IUCS2009*, ACM, pp. 30–35, 2009.
19) Miyabe, M,, Yoshino, T. and Shigenobu, T., "Effects of repair support agent for accurate multilingual communication," in *PRICAI 2008: Trends in Artificial Intelligence*, *LNCS 5351*, Springer Berlin / Heidelberg, pp. 1022–1027, 2008.
20) Morita, D. and Ishida, T., "Collaborative translation by monolinguals with machine translators," in *Proc. of the 14th IUI2009*, ACM, pp. 361–366, 2009.
21) Morita, D. and Ishida, T., "Designing protocols for collaborative translation," in *PRIMA '09, PRIMA '09*, Berlin, Heidelberg, pp. 17–32, 2009.
22) Pascual-Nieto, I., Santos, O. C., Perez-Marin, D. and Boticario, J. G., "Extending computer assisted assessment systems with natural language processing, user modeling, and recommendations based on human computer interaction and data mining," in *Proc. of the IJCAI'11*, pp. 2519–2524, 2011.
23) Popović, M. and Ney, H., "Towards automatic error analysis of machine translation output," *Comput. Linguist.*, *37*, *4*, pp. 657–688, December 2011.
24) Pym, P. J., "Pre-editing and the use of simplified writing for mt: an engineer's experience of operating an mt system," in *Translating and the computer*, pp. 80–96, London, 1990.
25) Rafaeli, S., "Interactivity: From new media to communication," *Sage Annual Review of Communication Research: Advancing Communication Science*, *16*, pp. 110–134, 1988.
26) Schwier, R. A. and Misanchuk, E. R., "Interactive multimedia instruction," in *Englewood Cliffs, NJ: Educational Technology Publications*, pp. 19–33 1993.
27) Siddharthan, A. and McKeown, K., "Improving multilingual summarization: using redundancy in the input to correct mt errors," in *Proc. of the HLT '05*, pp. 33–40, 2005.
28) Song, W., Finch, A. M., Tanaka-Ishii, K., and Sumita, E., "picotrans: an icon-driven user interface for machine translation on mobile devices," in *Proc. of the 16th IUI2011*, ACM, pp. 23–32, 2011.

29) Specia, L., Raj, D. and Turchi, M., "Machine translation evaluation versus quality estimation," *Machine Translation*, *24*, pp. 39–50, 2010.
30) Swift, J. S., "Foreign language ability and international marketing," *European Journal of Marketing*, *25*, *12*, pp. 36–49, 1991.
31) Tanaka, S., Nakakubo, A., Kimura, M., Takeda, K. and Koda, T., "chapter Development of a multilingual translation service for interpretations and usage examples of mobile phone pictograms," *Culture and computing*, pp. 113–126, 2010.
32) Wilks, Y., *Machine Translation: Its Scope and Limits*, Springer, 2009.
33) Wu, H., Wang, H. and Zong, C., "Domain adaptation for statistical machine translation with domain dictionary and monolingual corpora," in *COLING*, pp. 993–1000, 2008.
34) Yamashita, N. and Ishida, T., "Automatic prediction of misconceptions in multilingual computer-mediated communication" in *IUI06*, 2006.
35) Yamashita, N. and Ishida, T., "Effects of machine translation on collaborative work," in *Proc of the 20th CSCW2006*, ACM, pp. 515–524, 2006.



Chunqi Shi, Ph.D.: He received his M.E. degree from the Institute of Computing Technology of the Chinese Academy of Sciences in 2007, subsequently became an engineer at NeatEase Inc. and Baidu Inc. until 2010, and earned his Ph.D. in informatics degree from Kyoto University in 2013. He is currently a postdoctoral researcher of the Department of Computer Science in Brandeis University. His achievements about user-centered design of translation systems have been published at international conferences such as the ICWS or IUI. His research interests lie in services computing, intelligent interfaces and multi-lingual system design.

Toru Ishida: He is a professor of Department of Social Informatics and Design School, Kyoto University. He has been a fellow of IEEE, IPSJ, and IEICE since 2002, 2005, and 2008. His research interest lies with autonomous agents and multi-agent systems. He served as a program co-chair of the second ICMAS, and a general co-chair of the first AAMAS. He was an editor/associate editor of Journal on Web Semantics, IEEE PAMI and Journal on Autonomous Agents and Multi-Agent Systems. He started workshop/conference on Digital Cities and Intercultural Collaboration. His current projects including the Language Grid, a multilingual service infrastructure on the Internet.

Donghui Lin, Ph.D.: He received his M.E. degree in computer science and engineering at Shanghai Jiao Tong University in 2005, and Ph.D. degree in social informatics at Kyoto University in 2008. During 2008 to 2011, he was a researcher of National Institute of Information and Communications Technology, Japan. He is currently an assistant professor of Department of Social Informatics at Kyoto University. His research interests include services computing, services science, business process management, crowdsourcing and intercultural collaboration. He served as a program co-chair of Culture and Computing 2013, and a program committee member/session chair of recent major international conferences in the area of services computing and services science including IEEE SCC, IEEE ICWS, IEEE SOSE and CLOSER.