The Chinese Room: Visualization and Interaction to Understand and Correct Ambiguous Machine Translation

Joshua Albrecht[1], Rebecca Hwa[1], G. Elisabeta Marai[1]

[1]University of Pittsburgh, Department of Computer Science

Abstract

We present The Chinese Room*, a visualization interface that allows users to explore and interact with a multitude of linguistic resources in order to decode and correct poor machine translations. The target users of* The Chinese Room *are not bilingual and are not familiar with machine translation technologies. We investigate the ability of our system to assist such users in decoding and correcting faulty machine translations. We found that by collaborating with our application, end-users can overcome many difficult translation errors and disambiguate translated passages that were otherwise baffling. We also examine the utility of our system to machine translation researchers. Anecdotal evidence suggests that* The Chinese Room *can help such researchers develop better machine translation systems.*

Categories and Subject Descriptors (according to ACM CCS): I.3.3 [Computer Graphics]: Information Visualization—Machine Translation

1. Introduction

The field of Machine Translation (MT) is concerned with developing methods for automating the task of translating between two natural human languages, such as Chinese and English. However, because this task is difficult even for skilled human translators, and requires a considerable amount of world knowledge that cannot be easily encoded in straightforward algorithms, the sentences produced by current MT systems are often difficult or impossible to understand. Consider the following example output:

```
"He utter eyes and not the
slightest attention As leakage."
```

The resulting output is more accurately described as a jumble of words than an English sentence, even though the output was produced by one of the best MT systems freely available today [Goo08,NIS06]. The translation for the original sentence should have been:

```
"His eyes were wide apart; noth-
ing in their field of vision es-
caped."
```

Many researchers are working hard to improve MT directly by creating better algorithms and systems. We pursue an alternative solution – allow human users access to the significant amount of information available to an MT system, and let the user correct the MT output. This idea has been proposed as far back as 1980 by Martin Kay [Kay80], but that work and subsequent approaches have focused on improving the performance of professional translators. Our goal is to allow even users who are not bilingual to gain most of the information contained in the original source sentence.

In our approach, the human user relies on the machine to do the "symbol-pushing," while the machine relies on the human's world knowledge to guide the search for a correct translation. The assumption is that, although our intended users do not know the source language and may not be familiar with MT technologies, they have enough world knowledge and linguistic abilities in their native language to help them "decode" the disfluent MT output.

The process of language translation is complex and depends on different types of linguistic information; some of these information sources may contradict each other. Access to all this information could quickly overwhelm a potential

user, so it is essential that we design an effective system that presents the information visually, in a useful and understandable manner.

We present a prototype of a collaborative translation system, called *The Chinese Room*. The system visualizes ambiguous linguistic information about the unknown foreign language as it relates to the user's native language so that the user may gain an intuitive feel for what the source text might mean and thus overcome the mistakes made by the MT system.

Moreover, to better understand how users interact with the system and help MT researchers design stronger automated translation systems, we have designed an analysis module based on the timeline work of Plaisant et al. [PMR*96]. A contribution of this work is that we demonstrate a novel domain application in which visualization helps machine translation research.

2. Background and Related Work

2.1. Background

In the process of translation, typical MT systems use a variety of linguistic resources — in terms of both data and tools. Each of these tools and data sources — briefly reviewed below — may introduce errors in the translation process.

In a first step, the source text is typically *segmented* into words; the task is not trivial for Chinese text, which contains no spaces naturally. As with all such automatic linguistic tools, errors are made during the segmentation, which can cause further errors later in the MT process.

Each sentence in the source text can also be *tagged* and *parsed*, in order to test conformability of the sentence to a logical grammar. The tagging step labels each word with a part-of-speech (POS) tag, such as noun, verb, or adjective. The POS-tagged sentence is then parsed; the parsing step produces a parse of the sentence — a tree structure showing the relationships among words within the sentence. Errors in the tagging and parsing steps can also propagate to the MT output.

Using the parsing information and *dictionaries*, the words are finally translated to the target language. Additional errors are possible at this stage. For example, the Chinese character pronounced *mei3* could mean either the noun "beautiful", or the noun "the United States." Choosing one definition over the other within a sentence can lead to very different machine translations: *He was responsive to beauty...*, as opposed to *He was sensitive to the United States....*

The output of the MT system is not only the translated sentence, but also a mapping between the words in each language. This mapping is called the *alignment*. In addition to the translation and alignment, some MT systems can also provide the *n-best* retranslations of a group of words. These retranslations are alternative translations generated by the MT system, sorted in decreasing order of what the MT system considers to be their probability of being correct.

In addition to bilingual dictionaries, an MT system may consult resources such as *glosses*, *monolingual* or *bilingual corpora*. A gloss is a brief summary of a word's meaning, equivalent to the dictionary entry of that word, but only a word or two in length; it serves as a simple translation. A corpus (plural corpora) or text corpus is a large and structured set of texts typically used for statistical analysis. In more recent approaches, MT systems may also search the web for already-translated similar phrases.

2.2. Related Work

While there is a considerable body of knowledge on the design of MT systems, there has been surprisingly little work in the area of visualizing machine translation. Recently, a tool called DerivTool [DKC05] was created for the purposes of interacting with the core of an MT system. However, DerivTool's focus was on directly improving a specific MT system, and as such required in-depth knowledge of the particular MT system under analysis. Our application handles generic MT data and targets users who are not familiar with MT technologies.

The idea of leveraging human-computer collaborations to improve MT is not new; computer-aided translation, for instance, was proposed by Kay [Kay80]; research systems and commercial products have been successfully developed [Bow02,LFL00]. The focus of these efforts has been on improving the performance of professional translators. In contrast, *The Chinese Room* is targeted at users who cannot read the source text. Our objective is also related to that of cross-language information retrieval [ROL01] in that we want to help users to gain a deeper understanding of the information in the documents retrieved.

The name of our system was inspired by Searle's *Chinese Room* thought experiment [Sea80], although there are major differences between our system and Searle's description. Most notably, our users manipulate Chinese symbols by inserting their knowledge rather than purely operating based on instructions; nonetheless, the name was evocative in that our users require additional resources to process the input symbols.

Established methods exist for visualizing and interacting with the various components — from text to trees — that come into play in our machine translation application. Our overall design of *The Chinese Room* is based on the graphical design principles outlined by Tufte [Tuf90] and on the classic interaction principles of Card et al. [CMS99]. Our visualization and browsing of parse trees was further inspired by the work of Munzner et al. [MGT*03].

Finally, the analysis module we designed in order to investigate how users collaborate with *The Chinese Room* builds on the timeline work of Plaisant et al. [PMR*96]
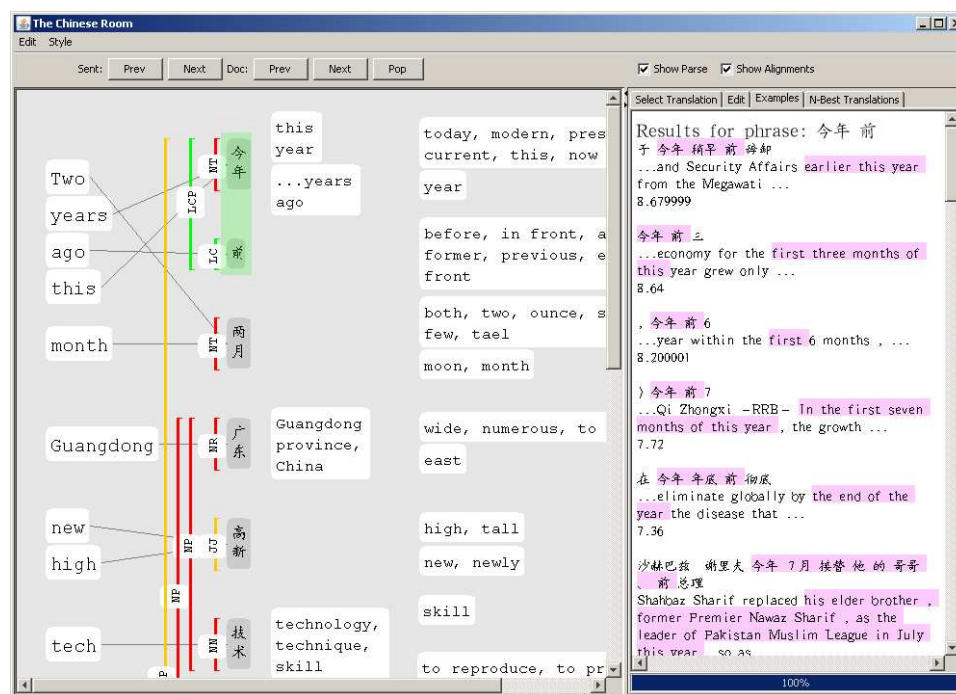
Figure 1: *The graphical environment consists of two main panes. The left pane is a workspace for exploring the sentence, while the right pane consists of multiple tabs that provide additional functionalities. The workspace displays the machine translation (leftmost column), source sentence, and alignments between them; the source sentence is annotated with its parse tree (colored brackets), its word and character glosses (two right columns). Currently displayed in the right pane is the example tab, showing the search results (highlighted in pink) for the selected Chinese phrase (highlighted in the left pane in green).*

3. Methods

3.1. Overview

In designing the prototype for *The Chinese Room*, we attempt to present the users with as many of the resources commonly used by MT systems as possible. Although many language-processing tools and multilingual resources are available as off-the-shelf packages, most are still imperfect. Finding the optimal way to integrate and display the possibly conflicting information from these resources is a challenging problem.

The Chinese Room consists of a visualization interface interconnected with five off-the-shelf text-processing modules: a machine translation and alignment module (the research version of Google's free Machine Translation service [Goo08]), a part-of-speech tagger (POS-tagger) and a parser module [KM03], a segmentation module (a by-product of Google's translation process), a custom glosses builder (we used the Chinese-English Translation Lexicon released by the Linguistics Data Consortium), and a custom information retrieval engine [Lem06], which allows the users to search large monolingual and bilingual corpora for approximate matches to difficult phrases. For corpora, we used the Federal Broadcast Information Service corpus and the Chinese Gigaword corpus. We chose five widely used resources that were freely available and had a reasonably good performance. Additional less-common resources — such as translation phrase dictionaries, multiple dictionaries, or translation rules for phrases — could be added to the application; exploring the performance of alternative translation algorithms and additional resources in the context of our system goes beyond the scope of this paper. Each text-processing module provides linguistic information, while the visualization interface presents these linguistics resources to the users in an intuitive fashion, and also facilitates the user interaction.

An additional behind-the-scenes module (Section 3.4) allows the MT researchers to examine visually the MT corrections performed by users. The module is packaged with the application, but it is typically accessed only by the MT researchers.

3.2. The Chinese Room Visual Display

Once the five text-processing modules process the source text, the resulting data must be displayed to the user in a way that is both simple enough to understand, and comprehensive enough for the task of understanding and correcting the
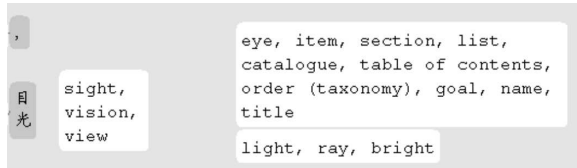
Figure 2: *Source text segmentation: the ideograms for 'eye' and 'light' or 'ray' form the word 'sight'.*

Figure 3: *Source to target alignment showing the MT mappings between source words in Arabic and target words in English. POS-tags label the parse tree, although novice users tend to focus on the tree structure and at most the first letter of each label: NP (noun phrase), NN (single noun), VBD (verb past tense), DT (determiner) etc.*

translation to be possible. The graphical display design was guided by iterative prototyping and feedback from both MT researchers and novice users. Early prototypes attempted to show the user simultaneous multiple views of the resources (for example, the document view, the sentence view, and the detail view — such as the alternative translations). However, feedback indicated that showing more than two views at a time was confusing the users. Later, more successful designs emphasized instead the translation task by showing the current translated sentence at all times in a left pane, while allowing the user to interactively view either details on demand or the document context in a second pane (Fig. 1). The left pane serves as a large workspace in which the user can interact with the translated text sentence by sentence; on the right pane are tabbed panels that give users access to information with additional context.

The left pane combines five sources of information – the most important and easier to use according to early feedback: the segmented source sentence, its translation, the alignment of the source and target sentence, the parse structure of the sentence, glosses for words and, in the Chinese case, glosses for characters. The text for both the initial machine translation and the dictionary definitions — indicated by user feedback as convenient resources — is displayed clearly in the white, rounded boxes. Text for the source sentence is shown in darker grey boxes, giving it less visual prominence because Chinese or Arabic characters are not directly useful to our users, who can't read them. The segmentation, which was useful to our users, is still readily apparent. For example, in Fig. 2 the users can see how the MT system built the word 'sight' from two ideograms, 'eye' and 'light'.

Alignments between the source words and the target words (Fig. 3) are shown in a dark grey color because they are often wrong or uninformative. The alignments allow the users to visually detect potential misalignments or poor word reordering. For instance, the automatic translation shown in Figure 1 begins: *Two years ago this month...* It is fluent but incorrect. The crossed alignments offer users a clue that "two" and "months" should not have been split up. English words in the machine translation are clustered together based on these alignments. The intuition is that alignments group words into likely phrasal units of translation, making it eas-
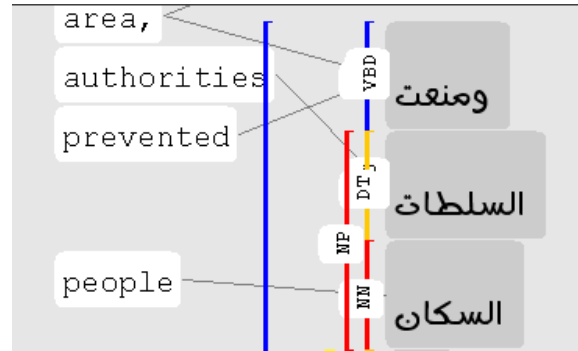
ier to see at a glance (a) how the sentence is structured, and (b) if any phrase looks out of place.

Glosses for words and characters are shown on the right side of the workspace pane. In the case of Chinese text, the placement of the word glosses presents a challenge because there are often alternative Chinese segmentations. We place glosses for multi-character words in the column closer to the source. When the user mouses over each definition, the corresponding characters are highlighted, helping the user to notice potential mis-segmentation in the Chinese. Dictionary definitions are organized based on the characters that they correspond to.

Finally, the source sentence is annotated with its parse structure. The design of the parse tree was challenging, since most users are not familiar with the concept of a parse tree. In such a tree, each node in the tree represents either a syntactic phrase (if it is an interior node), or a POS-tagged word (if it is a leaf node). We represent each syntactic node with a bracketed line that spans each of the child phrases and words. The brackets are color-coded into four major types (noun phrase, verb phrases, prepositional phrases, and other). Each node is also labeled with the name of the phrase so that the mapping between color and type does not need to be remembered by the users. Early feedback indicated the parse structure was useful in general when analyzing a phrase, while POS-labels were more useful to MT researchers than to novice users. Accordingly, in our pilot study the users were instructed to focus on the extent of brackets rather than their color-mapping and labels; future versions of the system will enable the users to turn off POS-labeling.

The right pane of the visual display can be used (1) as an overview tab that shows all sentences in the document; sen-

tences can be selected interactively to be worked on in the left pane; (2) as a notepad-like tab that allows direct editing of the machine translation; or (3) to display additional information (details-on-demand) such as alternative translations or similar phrases.

3.3. Interaction

When users encounter a problem that requires more information than is displayed by default, they have two options for exploration — searching for similar phrases, or requesting n-best retranslations for that phrase. Users can select a portion of the source string in order to search for similar phrases in a bilingual corpus; the search returns professional translations in similar contexts. Additionally, phrases from a large monolingual Chinese corpus are also returned together with their automatic translations. If the users wish to examine any of the translation pairs in detail, they can push it onto the sentence workspace. Finally, users can also request alternative translations from the MT system for selected sources phrases. The *N*-best alternatives are displayed.

In the left pane, users can zoom-in, collapse and expand the parse-tree brackets to keep the workspace uncluttered as they work through the source sentence. This action also indicates to the MT researcher which fragments held the user's focus. Highlighting either the original Chinese word or the definition in the left pane will show the matching definition or characters, respectively. Finally, all English elements can be edited and dragged around the screen, allowing the user to tangibly interact with the sentence, and consider alternative target-word orderings and thus alternative translation possibilities.

3.4. Visual Analysis

One of the issues of prime interest to the MT researchers was investigating how exactly people use *The Chinese Room*. What resources and strategies did they use? To answer such questions, and in collaboration with the MT researchers, we created a timeline visualization for each trial; a trial is defined as one document being worked on by a single user. The timeline diagrams provide a nice, simple description of the user behavior during the trial.

The analysis window (Fig. 4) is split horizontally into two main white panes. The MT researcher can load one trial — i.e., document/user pair — in each pane. The two panes allow for easier comparison of the trial pairs.

Each pane shows the user's actions for a particular document. Vertical lines of color represent the user actions on a specific sentence within a document; each horizontal line corresponds to a sentence within the document, and the x-axis is mapped to time. Each color represents a different action: editing, document context view, search results, alternate translation, or *other* — for example, examining an

example sentence in detail. Thus, colors represent which resources were used, and when.

The analysis window also contains two toggle buttons. The 'Type' button controls whether the plot boxes are showing cumulative action or the timeline of the actions. The 'Connected' button controls whether the time scale of the two boxes is scaled separately or is locked together. We note that the analysis module was designed with the goal of comparing the behavior of a few users at a time, which was a good fit with the exploratory nature of the present study. The module would likely have to be revised to fit potential larger-scale studies.

4. Evaluation and Results

To evaluate the MT impact of *The Chinese Room*, we have conducted a pilot experiment. We asked eight non-Chinese speakers to correct the machine translations of four short Chinese passages, approximately ten sentences long each. While both the participant pool and the dataset (184 participant-corrected sentences) are relatively small, they allowed the MT researchers to perform quantitative and qualitative assessments while controlling for user backgrounds and experiences (see [AHM09] for a detailed analysis).

Each participant was instructed to (a) correct the translations for one news article and one fiction passage using all the resources made available by *The Chinese Room* and (b) correct the other two passages without *The Chinese Room*. To keep the experimental conditions as similar as possible, for task (b) we provided the users with a restricted version of the graphical interface (Fig. 5) in which all additional functionalities except for the Document View Tab were disabled. Half of the users began with task (b), and the other half with task (a). Thus, every passage received four sets of corrections made collaboratively with the system, and four sets of corrections made based solely on the participants' internal language models.

The participants were asked to complete each passage within one session, without further time constraints. Within a passage, the users could work on the sentences in any arbitrary order. They could also elect to "pass" any part of a sentence if they found it too difficult to correct. Timing statistics were automatically collected. We conducted a short exit interview with each participant at the end of the session.

The corrected translations were evaluated by two bilingual speakers ("judges"). The judges were presented with the original source text as well as the parallel English text for reference. Each judge was then shown a set of candidate translations: the original MT output, an alternative translation by a bilingual speaker, and corrected translations by the participants, in a randomized order. Since the human corrected translations are likely to be fluent, we have instructed the judges to concentrate more on the adequacy of the meaning conveyed. They were asked to rate each sentence on an
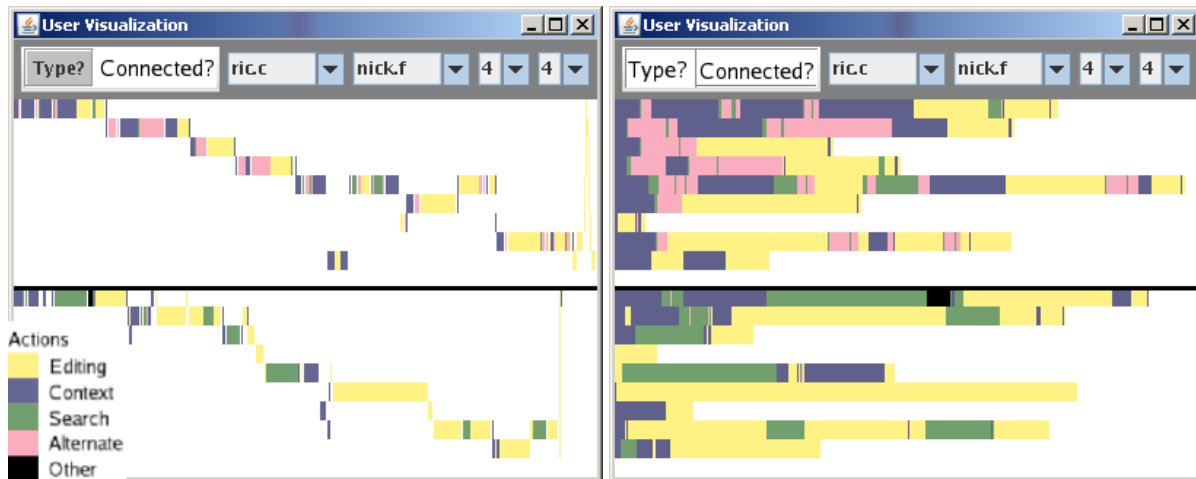
Figure 4: *Two views of the visual analysis window, showing the actions of two users (ric and nick) as they worked on the same document 4, both using the Chinese Room. Left: action timeline view: right: cumulative action view. The MT researcher quickly saw that the first user (ric) used a wide variety of resources, while the second user relied almost exclusively on searching for similar examples. Both users took a similar approach of working their way through each example sentence one by one without skipping around much, and relying instead on local context.*
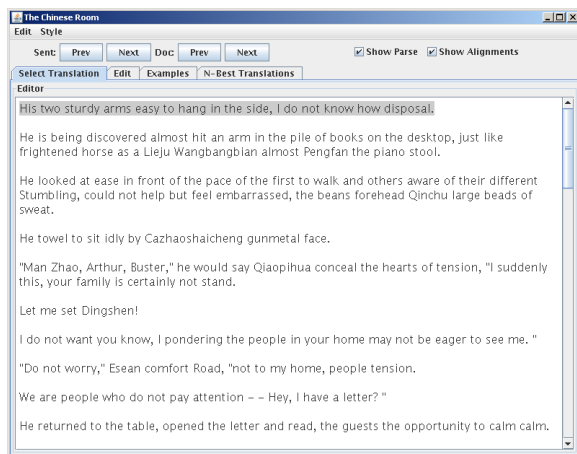

Figure 5: *The interface for users who are correcting translations without the Chinese Room; they have access to the document view, but they do not have access to any of the other resources.*

absolute scale of 1-10, where 9-10 means "The meaning of the Chinese sentence is fully conveyed in the translation", and 1-2 means "The translation makes no sense at all." To reduce the biases in the rating scales of different judges, we normalized the judges' scores, following standard practices in MT evaluation [BFF*03]; the normalization resulted in scoring in the [0,1] range.

Using *The Chinese Room*, the experiment participants were able to improve on average the MT quality from 0.35 to 0.53, closing the gap between the MT and bilingual translations by 36.9% without knowing the source language (the average score of bilingual translations was 0.83). Table 1 shows example outputs from the participants.

Overall, the MT outputs contained enough errors that the participants were able to improve, to a small degree, the MT quality even without *The Chinese Room*, from 0.35 to 0.42. These differences are all statistically significant (using a paired t-test with >98% confidence). In general, participants who used *The Chinese Room* had more instances of large improvements than participants who made corrections without help.

In many cases, multiple users were able to identify a translation error without being able to fix it, but one or two users managed to understand the intended meaning by working with our system. As an upper-bound for the effectiveness of the system, we construct a combined "oracle" user out of all 4 users that used the interface for each sentence. The oracle user's average score is 0.70; in contrast, an oracle of users who did not use the system is 0.54 (relative to the MT's overall of 0.353 and the bilingual translator's overall of 0.833). This suggests *The Chinese Room* affords a potential for human-human collaboration as well.

The higher quality of corrections did require the participants to put in more time. Overall, the participants took 2.5 times as long when they had access to *The Chinese Room* than when they did not. This may be partly because the participants have more sources of information to explore, and partly because the participants tended to "pass" on fewer

	Score	Translation
MT	0.336	He is being discovered almost hit an arm in the pile of books on the desktop, just like frightened horse as a Lieju Wangbangbian almost Pengfan the piano stool.
Without The Chinese Room	0.263	Startled, he almost knocked over a pile of book on his desk, just like a frightened horse as a Lieju Wangbangbian almost Pengfan the piano stool.
With The Chinese Room	0.778	He was nervous, and when one of his arms nearly hit a stack of books on the desktop, he startled like a horse, falling back and almost knocking over the piano stool.
Bilingual Translator	0.934	Feeling nervous, he discovered that one of his arms almost hit the pile of books on the table. Like a frightened horse, he stumbled aside, almost turning over a piano stool.

Table 1: *Example translation corrected by the participants and their scores. In this example, the initial MT was badly jumbled, but the informed user was able to recover most of the meaning.*

sentences. We note that the goal of this project is to improve the quality of machine translations, and not to minimize the task completion time.

During the exit interviews, the participants were asked: (a) to give an overall summary of each translated document; (b) about their overall satisfaction with the tool, including suggestions for improvement; (c) about the specific strategies they used to collaborate with the system. In general, summaries for news articles were accurate, while, unsurprisingly, summaries for story excerpts were less so (for example, one participant mistook a fragment from *Martin Eden* for a spy story). Participant experiences, as revealed through the exit interviews, were generally positive. Because the users felt like they understood the translations better, they did not mind putting in the time to collaborate with the system. Specific comments included: "happy to have it [the tool]", "at least it gives one something to do when one is stuck", "it was fun", while the suggestions for improvement caught several minor bugs in the user interface. Novice users did not find the POS-tags particularly useful, although having access to the parse tree structure was considered helpful.

During the exit interviews, the participants were also asked to describe strategies that they developed for collaborating with the system. Their responses fall into three main categories:

- **Divide and Conquer**: Some users found the parse trees helpful in identifying phrasal units, for which they subsequently required *N*-best retranslations or example searches. For longer sentences, they used the constituent collapse feature to help them reduce clutter and focus on a portion of the sentence.
- **Example Retrieval**: Using the search interface, users examined the highlighted query terms to determine whether the MT system made any segmentation errors. Sometimes, they used the examples to arbitrate whether they should trust any of the dictionary glosses or the MT's lexical choices. Typically, though, they did not attempt to inspect the example translations in detail.
- **Document Coherence and Word Glosses**: Users often referred to the document view to determine the context for the sentence they are editing. Together with the word

glosses and other resources, the context clues helped the users make better lexical choices than when they made corrections without the full system and relied on document coherence alone.

In general, users often accessed the document context view and requested alternative translations. About half as often they searched for similar examples and expanded or collapsed the right pane tree. The option of inspecting retrieved examples in detail (i.e., bring them up on the sentence workspace) was rarely used, perhaps because novice users experienced a greater degree of uncertainty than professional translators.

Figure 4 shows anecdotal evidence of how the visual analysis interface was useful to the MT researchers: two users are shown correcting the same document, both using *The Chinese Room*. In this particular example, the MT researcher can quickly see that the first user (ric) used a wide variety of resources, while the second user (nick) relied almost exclusively on searching for similar examples. Both users took a similar approach of working their way through each example sentence one by one, in a sequence, and relying primarily on the local context. The MT researcher also noticed that both users cared enough to put some effort into editing the final pass, even though this was the last document and they had been using the *Chinese Room* for several hours; in Fig 4, see the little "editing pass" at the end of the trial, where the users ensured that each sentence was what they wanted. The researcher noted: "If we were to look back at previous documents done by these users, I could see how their approach changed over time."

In general, the most important thing the MT researchers learned from the analysis visualizations was a sense for how different each person's approach to the problem was. As confirmed by the exit interviews, some people skipped all over, some relied on certain resources much more than others, while some would spend half the time on the first sentence, then do the other sentences very quickly. The researchers had hoped to be able to say more about the usage of various linguistic resources, but since the pilot users used the interface in very different ways, it was difficult to make generalizations. Perhaps with more detailed logging, or a larger

user pool, more commonalities could have been found. The researchers wished that we had recorded slightly more information, such as mouse clicks and movements, but on the other hand, they also felt they already had a lot of information about the user behavior.

5. Discussion and Conclusion

In this paper, we have shown that concepts from visualization and human computer interaction can have significant positive impacts on research in machine translation.

From our experiments, *The Chinese Room* seems promising as an end-application. Although the participants did not find all the translation errors using our current system, the corrections they made led to a significant improvement over the original machine translation. Moreover, our timeline analysis module helped researchers to understand what resources humans need to correct erroneous machine translations and how they interact with these resources to achieve this task. In the future, we plan to incorporate additional resources specific to a particular type of MT systems to help researchers further improve their systems. The timeline analysis may also lead to alternative interface designs. For example, although feedback from earlier prototypes indicates that simultaneous multiple views are confusing to the users, the timeline analysis shows that participants frequently switch between different views. Ultimately, the *Chinese Room* serves as a useful diagnostic tool to help designers of MT systems to verify ideas about what types of resources and data are useful for automatic translation.

One of the most interesting challenges in this work was finding ways to incorporate into the visual display the uncertainty inherent to the various machine translation subprocesses. In visualization, uncertainty is often ignored or treated as a binary quality. In knowledge discovery applications it is, however, important to show to the user not only that the value of a particular option is uncertain, but also the other options available. In some instances, by showing alternative options, our visualization enabled the users to unravel and repair sequences of three or four consecutive mistakes.

In conclusion, we have designed, implemented and tested *The Chinese Room*, a visualization interface that makes available a variety of linguistic resources to users in an intuitive display and facilitates their collaborations with these resources. By combining evidence from complementary information sources, users can infer alternative hypotheses based on their world knowledge and improve the overall translation. Experimental evidence suggests that the collaborative effort between human users and our system results in much improved translations than either the original MT or uninformed human edits. Moreover, in several instances, the quality of the corrected translation approached that of bilingual speakers. User inputs may be gathered for error analysis and system training for future MT development. Finally, this work demonstrates a novel domain application in which visualization and interaction help machine translation research.

References

[AHM09] ALBRECHT J., HWA R., MARAI G. E.: Correcting automatic translations through collaborations between mt and monolingual target-language users. In *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics (in press)* (2009). 5

[BFF*03] BLATZ J., FITZGERALD E., FOSTER G., GANDRABUR S., GOUTTE C., KULESZA A., SANCHIS A., UEFFING N.: *Confidence estimation for machine translation*. Tech. Rep. Natural Language Engineering Workshop Final Report, Johns Hopkins University, 2003. 6

[Bow02] BOWKER L.: *Computer-Aided Translation Technology*. University of Ottawa Press, Ottawa, Canada, 2002. 2

[CMS99] CARD S., MACKINLAY J., SHNEIDERMAN B.: *Readings in information visualization: Using vision to think*. Morgan Kauffman, San Francisco, 1999. 2

[DKC05] DENEEFE S., KNIGHT K., CHAN H. H.: Interactively exploring a machine translation model. In *Proceedings of the ACL Interactive Poster and Demonstration Sessions* (Ann Arbor, Michigan, June 2005), pp. 97–100. 2

[Goo08] GOOGLE: Google machine translation service. http://code.google.com/apis/translate/research. 1, 3

[Kay80] KAY M.: *The proper place of men and machines in language translation*. Tech. Rep. CSL-80-11, Xerox, 1980. Later reprinted in *Machine Translation*, vol. 12 no.(1-2), 1997. 1, 2

[KM03] KLEIN D., MANNING C. D.: Fast exact inference with a factored model for natural language parsing. *Advances in Neural Information Processing Systems 15* (2003). 3

[Lem06] LEMUR: Lemur toolkit for language modeling and information retrieval, 2006. The Lemur Project is a collaborative project between CMU and UMASS. 3

[LFL00] LANGLAIS P., FOSTER G., LAPALME G.: Transtype: a computer-aided translation typing system. In *Workshop on Embedded Machine Translation Systems* (May 2000), pp. 46–51. 2

[MGT*03] MUNZNER T., GUIMBRETIÈRE F., TASIRAN S., ZHANG L., ZHOU Y.: Treejuxtaposer: scalable tree comparison using focus+context with guaranteed visibility. *ACM Trans. Graph. 22*, 3 (2003), 453–462. 2

[NIS06] NIST: 2006 machine translation evaluation official results. http://www.itl.nist.gov/iad/mig/tests/mt/. 1

[PMR*96] PLAISANT C., MILASH B., ROSE A., WIDOFF S., SHNEIDERMAN B.: Lifelines: Visualizing personal histories. In *In Proceedings of ACM CHI 96 Conference on Human Factors in Computing Systems* (1996), ACM Press, pp. 221–227. 2

[ROL01] RESNIK P. S., OARD D. W., LEVOW G.-A.: Improved cross-language retrieval using backoff translation. In *Human Language Technology Conference (HLT-2001)* (2001). 2

[Sea80] SEARLE J.: Minds, brains, and programs. In *Behavioral and Brain Sciences* (1980), vol. 3:417, pp. 34–57. 2

[Tuf90] TUFTE E. R.: *Envisioning Information*. Graphics Press, Cheshire, Connecticut, USA, 1990. 2