# Using Bilingual Segments to Improve Interactive Machine Translation

Na Ye, Ping Xu, Chuang Wu, Guiping Zhang

Human-Computer Intelligence Research Center, Shenyang Aerospace University, Shenyang 110136, China
yn.yena@hotmail.com, 451908817@qq.com, {wuchuang,zgp}@ge-soft.com

**Abstract.** Recent research on machine translation has achieved substantial progress. However, the machine translation results are still not error-free, and need to be post-edited by a human translator (user) to produce correct translations. Interactive machine translation enhanced the human-computer collaboration through having human validate the longest correct prefix in the suggested translation. In this paper, we refine the interactivity protocol to provide more natural collaboration. Users are allowed to validate bilingual segments, which give more direct guidance to the decoder and more hints to the users. Besides, validating bilingual segments is easier than identifying correct segments from the incorrect translations. Experimental results with real users show that the new protocol improved the translation efficiency and translation quality on three Chinese-English translation tasks.

**Keywords:** Interactive Machine Translation, Bilingual Segment, Translating Option, Option Diversity.

## 1 Introduction

The performance of machine translation (MT) systems has been greatly improved by the statistical machine translation (SMT) and the neural machine translation (NMT) technology. However, in many tasks which have high quality requirements, the MT output is still not good enough and must be corrected by human translators in a post-editing (PE) stage.

To enhance the human-computer collaboration, Foster [1] introduced the interactive machine translation (IMT) technology. In an IMT system, a correction-prediction process works iteratively. First, the IMT system provides a raw translation. Second, the user validates the longest correct prefix in it and corrects the next word. Third, the system predicts a new suffix which is expected to be better than the previous one. This process is repeated until the correct translation is acquired.

During these years, the IMT technology was developed along with the evolution of the underlying MT models from SMT [2-6] to NMT [7-8]. Advances also include making better use of the prefix to improve the prediction accuracy [9-14], applying confidence measures to reduce human effort [15-16], adopting active learning [17] and online learning [18-19] to learn from user feedback, and integrating automatic speech recognition [20] and handwritten recognition [21] to multi-modal interaction tasks. Evaluation results show that compared to PE, the prefix-based IMT protocol can increase the human translation quality and reduce the number of key strokes while keeping the translation speed [22-24].

Recently, this left-to-right protocol was extended to make the human-computer interaction more flexible [7, 25]. In the extended protocol, users can validate the segments that should be kept in the translation. However, this protocol also suffers from three issues. First, the positions of the validated segments are not known, so the search process can only be constrained on a soft way [7]. Second, the user validations are restricted to the proposed translation, and no hints of other translating options are available. Third, identifying correct segments from incorrect translations often requires considerable cognitive effort, especially when the translation quality is low.

In this paper we refined the interactivity protocol to validating bilingual segments. During interaction, users are provided with both the source segments and their corresponding translating options. They can select the correct one from these options. In the new protocol, the target-side segments are aligned with their source-side counterparts, so they can be introduced to the decoder on a more direct way. This protocol also provides more hints to the users and requires less cognitive burden. We conducted experimentation with real users on three Chinese-English LDC corpora featuring different domains. Results show significant improvements over the prefix-based system.

## 2 Interactive Machine Translation

### 2.1 Prefix-based IMT

In the traditional prefix-based IMT protocol [2], the system predicts the best suffix under the condition of the given source text $s$ and the user-validated prefix $t_p$ as follows:

$$\hat{t}_s = \operatorname*{argmax}_{t_s} P(t_s \mid s, t_p) = \operatorname*{argmax}_{t_s} P(t_p, t_s \mid s) \tag{1}$$

where $(t_p, t_s)=t$, indicating that the prefix $t_p$ and the predicted suffix $t_s$ concatenate to form a complete translation $t$. To model $P(t_p, t_s \mid s)$, current approaches filter the translation hypotheses according to their matching results with the prefix.

Although the suffix is probably not quite right, there are some correct segments which should be kept. However, in the prefix-based IMT systems, these segments may lose in the next iteration. The segment-based IMT approaches were proposed to solve this problem.

### 2.2 Segment-based IMT

In this protocol [7], users can validate the segments $f_1^N = f_1, \ldots, f_N$ that should be retained in the future interactions. The search process can be modelled by predicting the other segments as follows:

$$\begin{aligned} \hat{g}_1^N &= \underset{g_1^N}{\operatorname{argmax}}\ P(g_1, \ldots, g_N \mid s, f_1, \ldots, f_N) \\ &= \underset{g_1^N}{\operatorname{argmax}}\ P(f_1, g_1, \ldots, f_N, g_N \mid s) \end{aligned} \tag{2}$$

where $g_1^N = g_1, \ldots, g_N$ is the *non-validated* segment sequence that fills $f_1^N$ to form a new translation. The prefix $t_p$ is a particular case of the validated segments. In Eq. (2) the search space is all possible hypotheses containing the segment set $\{f_1, f_2, \ldots, f_N\}$. Figure 1 shows an example of the segment-based IMT protocol.

| | |
|---|---|
| Source | 根据 本条 规则 发出 的 原诉 传票 须 采用 附录 A 表格 10 的 格式 。 |
| MT | The originating summons under section shall be in Form No. 10 in Appendix A. |
| Reference | An originating summons under this rule shall be in Form No. 10 in Appendix A. |

**Fig. 1.** Example of the Segment-based IMT Protocol.

In the above figure, framed texts indicate the validated segments. These segments are monolingual and lack the alignment information with the source sentence. The decoder can only predict the next words according to the previously generated words and the immediate next segment, and no source-target alignment information can be used. Thus the guidance of segments in decoding is limited. Another problem is the wrong word "*section*" and the missing word "*this*". Although the user found these mistakes, he/she have no other options to choose from. It is also difficult for users to identify the correct segments from the MT output mixed with wrong words and wrong orders. These problems are intended to be solved in the bilingual segment based IMT protocol.

### 2.3 Bilingual Segment based IMT

In this protocol, the source segments are aligned with their target counterparts. For each source segment, multiple translating options are provided. The user can validate bilingual segment pairs in the form of $\langle f_i, e_i \rangle$. The best translation is acquired as follows:

$$\begin{aligned} \hat{t} &= \underset{t}{\operatorname{argmax}}\ P(t \mid s, f_1, e_1, \ldots, f_N, e_N) \\ &= \underset{t}{\operatorname{argmax}}\ P(f_1, e_1, \ldots, f_N, e_N, t \mid s) \end{aligned} \tag{3}$$

where $e_i$ is the correct translation of $f_i$ validated by the user. The prefix $t_p$ is a particular case of segment pair which has no source counterpart. In Eq. (3) the search space is the hypotheses compatible with these bilingual segments. Figure 2 shows an example of the new protocol.

| | | |
|---|---|---|
| | Source | 根据 本条 规则 发出 的 原诉 传票 须 采用 附录 A 表格 10的 格式 。 |
| IT-1 | User | this rule an originating summons |
| | MT | Under this rule of an originating summons shall be in Form No. 10 in Appendix A. |
| IT-2 | User | A |
| | Accept | An originating summons under this rule shall be in Form No. 10 in Appendix A. |

**Fig. 2.** Example of the Bilingual Segment based IMT Protocol.

The user validated three bilingual segments (e.g. the framed parts in the figure). And the MT decoder proposed a translation better than that in Figure 1. Then the user inputs a prefix "*A*" and decodes again. This time the correct translation is acquired (IT-2).

## 3 User Interface

### 3.1 Overview

This paper adopts a user interface illustrated in Figure 3. The interface consists of two zones. One is the interacting zone which presents the segmented source sentence and the translating options. The segments and the options are aligned vertically by the left. When the mouse hangs over a source segment, a menu with its highest ranked options is displayed and the users can click the best option for validation. The other is the editing zone which presents the MT output when the users finish validation and press the "*translate*" button. Here the users can make corrections as they want, until the translation is accepted. The interacting and the editing process can alternate.

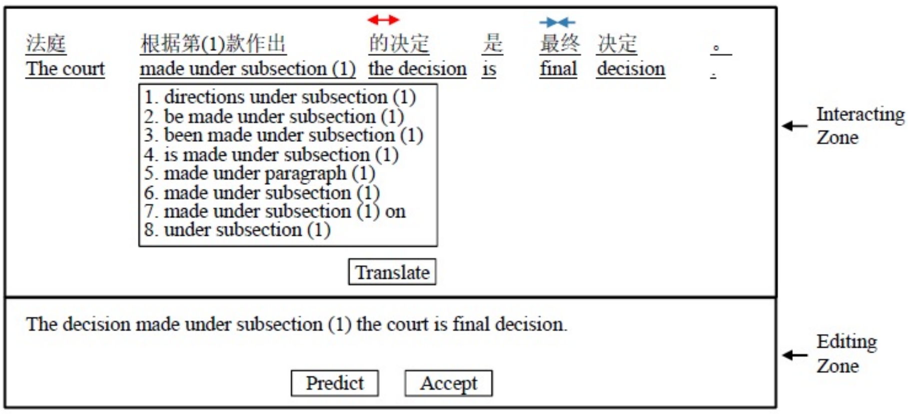


**Fig. 3.** User Interface of the Bilingual Segment based IMT System.

One prominent advantage of phrase-based SMT is the extraction for long phrases translation. Taking long phrases as the basic translating units can effectively alleviate the word disambiguation problem, thus achieving good results. Therefore, longer segments and their translations are displayed with priority in the interface, and the initial segmentation of the source sentence is performed by forward maximum matching algorithm with the phrase table. The displayed translating options are the highest $K$ options in the phrase table.

The interface also provides three auxiliary functions, namely segment split-merging, translating option re-ranking and suffix predicting.

### 3.2 Segment Split-merging

Above each segment, there are two kinds of two-head arrows. One is the splitting arrow (↔), it can split the segment into two smaller segments. The other is the merging arrow (→←), it can merge the segment and its next segment into one larger segment.

The arrows appear when the mouse hangs over the segment. If no smaller or no larger segment exists in the phrase table, then the arrow will not show. Once new segments are formed, their translating options will change accordingly.

### 3.3 Translating Option Re-ranking

By default, the options of a segment are ranked and displayed with their order in the phrase table. However, the highest scoring options are sometimes quite similar. So we offer an alternative mode that increases the diversity of options. The user can choose the default mode or the re-ranking mode before they start translation.

In this mode, the top $N$ translating options in the phrase table are re-ranked to generate a new option list. For each source phrase $p$, set a new option list $T$ ($T=\emptyset$). First add the option with highest score in the original phrase table to $T$. Then traverse the remaining $N$-1 options and find the one with highest diversity with the options in $T$ and insert it to $T$. Repeat this process until all the $N$ options are re-ranked. The diversity between two options $t_i$ and $t_j$ is measured as follows:

$$D\left(t_i, t_j\right) = 1 - \frac{c(t_i, t_j)}{\max\{|t_i|, |t_j|\}} \tag{4}$$

where $c(t_i, t_j)$ is the number of overlapped words (lemmatized) between $t_i$ and $t_j$.

### 3.4 Suffix Predicting

In the editing zone, the users can press the "*Prediction*" button to get the predicted suffix from the system. When the button is pressed, the current position of the cursor is recorded, and the characters before the cursor are taken as the prefix. The validated bilingual segments and the prefix are all considered to find the best suffix. Once a new suffix is generated, it will replace the current suffix. If the decoder fails to find any compatible hypothesis, then the suffix will not be changed.

# 4 Decoding

After the user finishes the validation of bilingual segments in the interacting zone, the system catches the user's choice of translating option for each segment $f_i$ and the current segmentation $S$ of the source sentence. A set is constructed as the constraints for decoding:

$$C = \{S, < p_1, f_1, e_1 >, < p_2, f_2, e_2 >, \dots, < p_N, f_N, e_N >\} \tag{5}$$

where $p_i$ is the position of segment $f_i$ in the source sentence. Recording $p_i$ is for avoiding the ambiguity when a segment appears more than once. Note that the user must click on the option before the segment and the option can be considered a validated bilingual segment. If a segment has never been clicked on any of its options, then the segment and its option cannot be used as a constraint.

We use the multi-stack-decoding algorithm for the phrase-based SMT decoder. Two improvements are proposed to meet the constraints. First, in order to make the generation of translation hypotheses compatible with the constraints, $S$ is used as the only segmentation of the source sentence during decoding. Second, the translating options of each source phrase (segment) are constrained by $<p_i, f_i, e_i>$, and only the options containing $e_i$ will be kept and participate in the subsequent decoding process.

As for the auxiliary function of suffix predicting, one more constraint is added to the decoder. The translation hypotheses must match the given prefix $t_p$.

# 5 Experiments

## 5.1 Data Setup

We tested the proposed approach on three different Chinese-English translation tasks with real users. *Laws* consists in laws texts from LDC2000T47 corpus. *Hansards* consists in Hansards texts from LDC2000T50 corpus. *News* consists in news texts from LDC2000T46 corpus. Table 1 gives the main figures of these corpora (S, T and V account for number of sentences, number of tokens and vocabulary size, respectively. k and M stand for thousands and millions).

**Table 1.** Main Figures of the Evaluation Corpora.

| | | Training | | | Development | | | Test | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | S | T | V | S | T | V | S | T | V |
| *Laws* | Zh | 103k | 2.0M | 29k | 2070 | 49k | 5.3k | 75 | 1.4k | 533 |
| | En | | 2.1M | 29k | | 51k | 5.2k | | 1.2k | 570 |
| *Hansards* | Zh | 351k | 7.8M | 75k | 2497 | 64k | 8.1k | 75 | 1.2k | 592 |
| | En | | 9.2M | 91k | | 78k | 11.3k | | 1.3k | 573 |
| *News* | Zh | 190k | 4.8M | 64k | 2512 | 72k | 10k | 75 | 1.2k | 667 |
| | En | | 5.4M | 69k | | 86k | 14k | | 1.3k | 623 |

The Chinese portions of these data were pre-processed by the ICTCLAS word segmenter[1], and the English portions were tokenized and lowercased. GIZA++ was used for training word alignment models. IRSTLM [26] was used for training 5-gram language models. Moses [27] was used for building phrase-based SMT models, which include 14 default features. MERT [28] was used for adjusting feature weights.

Three IMT systems were evaluated in our experiments. *Baseline* refers to the prefix-based system [2] using PBMT model. *BiSeg* refers to the proposed system without option re-ranking. *BiSeg+D* refers to the proposed system with option re-ranking. In the user interface, the number of displayed translating options is set to 10, and the number of top translating options for re-ranking is set to 20.

### 5.2 Evaluation Metrics

In the literature of IMT, automatic evaluation metrics are mostly used for assessing prototypes [2, 4, 7] because experiments with real users are quite costly and slow. In these metrics user behaviors are simulated, rather than real user behaviors during interaction. A direct evaluation of an IMT system would require conducting experiments with human users [7], which is done in this paper.

We evaluate the performance of an IMT system from two aspects of efficiency and quality. Translation efficiency is evaluated with three metrics:

1. *Translating Time*: is the most direct metric for measuring human effort during the IMT process. It is defined as the average time spent in translating each sentence.
2. *Key Stroke and Mouse-action Ratio* (KSMR) [2]: measures the user's keyboard and mouse effort during the IMT process.
3. *Decoding Times*: measures the number of decoding times during the IMT process. It is defined as the average decoding times for each sentence.

Translation Quality is evaluated with the *BLEU* score [29]. We take the English portions of the original parallel corpus as the reference and evaluate the users' translation quality. Note that the final translations accepted by the users are all correct, although not exactly the same as the references.

### 5.3 Participants and Procedures

A group of 9 postgraduates (6 females) from our research group volunteered to perform the evaluation as non-professional translators. All of them are native speakers of Chinese, and proficient in English.

In order to make the participants familiar with the systems, we selected 30 sentences in the *Laws* corpus as warm-up corpus, and let the participants translate with four IMT systems (10 sentences per system). Formal tests began on the second day after the warm-up.

[1] http://ictclas.nlpir.org/

We randomly divided the participants into 3 groups, 3 in each group. The testing set of each corpus was randomly divided into 3 parts, each with 25 sentences. The evaluation is carried out in a counterbalanced fashion.

### 5.4 Results and Analysis

In order to give an intuitive understanding of the performance of the SMT models, we evaluated the BLEU scores of the underlying SMT engine. Results are 0.3411, 0.1971 and 0.1901 on *Laws*, *Hansards* and *News*, respectively. This indicates that the quality of the translation provided by the SMT engine is readable with some effort.

Table 2 gives the average time of the three user groups on the evaluation corpora. The figures in the brackets are the relative differences between our systems and the Baseline system.

**Table 2.** Results of Translating Time with Different IMT Systems.

| | Baseline | BiSeg | BiSeg+D |
|---|---|---|---|
| *Laws* | 101.7s<br>(--) | 83.1s<br>(-18.3%) | 82.7s<br>(-18.7%) |
| *Hansards* | 103.0s<br>(--) | 79.5s<br>(-22.8%) | 78.6s<br>(-23.7%) |
| *News* | 91.5s<br>(--) | 86.1s<br>(-5.9%) | 84.2s<br>(-8.0%) |

It can be seen that the translating time with our systems is significantly lower than that with the Baseline system. This indicates a significant reduction in human effort. Through observing the IMT process of participants, we have the following findings:

1. The decoding times are less (see Table 4). Since there are many possibilities in segment translation and source sentence segmentation, it is difficult for the decoder of the prefix-based protocol to make the right choice. In the new protocol, the users not only validate the translation of segments, but also determined the segmentation of the source sentence. Thus the search space is greatly reduced.
2. The displayed option list provided by the new protocol is very helpful to the users, especially for long segments and in specialized domain. The segment-based protocol also does not provide this kind of help.
3. Under the new protocol, the users' IMT process can be divided into two stages, in which they focus on validating the segment translation and the segment order respectively. This two-stage translation process with clear task helps users to concentrate and improve their efficiency.

The option diversity can further reduce human effort. In most cases, the correct translating option is contained in the *K* displayed options. And the users don't always scan the options from top to bottom. In fact, as long as the correct option exists in the list, the users can easily pick it out. Especially for long segments, they usually have only a few options. However, there are also some cases (usually short segments) in which the correct option is not in the list. At this time, re-ranking the options increases the possibility of adding the correct option into the displayed list.

Table 3 gives the KSMR values on three corpora.

**Table 3.** Results of KSMR with Different IMT Systems.

| | Baseline | BiSeg | BiSeg+D |
|---|---|---|---|
| *Laws* | 0.94<br>(--) | 1.14<br>(+21.3%) | 0.99<br>(+5.3%) |
| *Hansards* | 0.55<br>(--) | 0.69<br>(+25.5%) | 0.66<br>(+20.0%) |
| *News* | 0.48<br>(--) | 0.71<br>(+47.9%) | 0.68<br>(+41.7%) |

We can see that the KSMR values of our systems are much higher than that of the Baseline system. The increased actions mainly come from four kinds of mouse actions, namely translating option clicking, segment splitting/merging, extra button clicking (such as the button "*translate*") and cursor switching between zones. But these mouse actions don't cost much thinking and action time, so they have little effect on translation efficiency.

Table 4 gives the average decoding times on three corpora.

**Table 4.** Results of Decoding Times with Different IMT Systems.

| | Baseline | BiSeg | BiSeg+D |
|---|---|---|---|
| *Laws* | 5.90<br>(--) | 4.52<br>(-23.4%) | 4.07<br>(-31.0%) |
| *Hansards* | 3.70<br>(--) | 1.89<br>(-48.9%) | 1.73<br>(-53.2%) |
| *News* | 2.80<br>(--) | 2.00<br>(-28.6%) | 1.97<br>(-29.6%) |

Table 4 shows that the decoding times reduce significantly in the new protocol. As mentioned above, the search space is greatly reduced. Through discussions with the participants, they report that it is better to validate the translation of content words and long segments and not to validate the function words and symbols. This avoids unnecessary constraints on decoding and reduces decoding failures.

Table 5 gives the translation quality (BLEU score) on three corpora.

**Table 5.** Results of Translation Quality with Different IMT Systems.

| | Baseline | BiSeg | BiSeg+D |
|---|---|---|---|
| *Laws* | 0.3811<br>(--) | 0.3846<br>(+0.9%) | 0.3834<br>(+0.6%) |
| *Hansards* | 0.2723<br>(--) | 0.2781<br>(+2.1%) | 0.2775<br>(+1.9%) |
| *News* | 0.2418<br>(--) | 0.2673<br>(+10.5%) | 0.2686<br>(+11.1%) |

Results show that the translation quality with our systems is slightly better than that with the Baseline system. This is because the new protocol provides a large number of options, giving users more hints and inspiration to enable them to produce better translations. While in the prefix-based protocol, the users will stop once they get an acceptable translation. However, the improvement is not obvious. The reason is that the users also work on good hypotheses under the prefix-based protocol. And the wrong translations are manually corrected before they can be submitted. So the users will spent more time but the translation quality can still be guaranteed.

Our protocol also has some disadvantages. For some sentences which the decoder can easily find the correct translation after a few word corrections, the prefix-based protocol is more convenient. While in our protocol the users still have to validate the translating options and then perform correction. We can use confidence scores to help users identify these cases and assign suitable protocol for each sentence.

### 5.5 Comparison with Related Work

Ananthakrishnan [30] proposed an interactive translation system which engages the user in a clarification dialogue to recover from the potential word-sense translation errors. This approach focuses on conversational spoken language translation and aims at solving the semantic class selection errors for certain ambiguous words. The users cannot perform additional interactive clarification operations.

Huang [31] proposed an input method for human translators. The MT technology is deeply integrated into the computer-aided translation system (CAT) to enable human translators to focus on choosing better translation results with less time. Experiments with real users show satisfactory results in accelerating the translation process. This work can be integrated into our protocol and further improve the IMT efficiency.

The work most similar to ours is that of Cheng [32] in which a pick-revise framework for IMT was proposed. This approach identifies the wrongly-translated phrase and selects the correct translation from the translation table. Then the sentence is re-translated with these constraints. In comparison, we extended the mathematical framework of the prefix-based protocol and the segment-based protocol through refining the constraints to bilingual segments. We also designed an interface for real users, which allows the users to split and merge the segmented phrases and provided the translating option re-ranking method that increases the option diversity. These help to improve the human efficiency in a real scenario. Besides, we conducted real-user experiments rather than simulated experiments.

## 6 Conclusion and Future Work

In this paper, we present a new IMT protocol. We provide users with translating options for each segment in the source sentence, and allow the users to validate bilingual segments in the user interface. In this way, more hints are given to the user and more direct guidance to the decoder. The protocol also helps users to concentrate in a two-stage translating process. We carried out real-user experiments on three corpora from different domains and obtained satisfactory results. Compared with the prefix-based IMT protocol, the new protocol made significant improvements in reducing human effort.

In the future, we will continue to optimize the user interface according to the users' suggestions. We will also consider applying the new input method to help users type faster and applying the confidence measures to determine the suitable protocol for the sentences. Integrating bilingual segments with the NMT framework is another problem worthy of study.

## Acknowledgements

This work is supported by the National Natural Science Foundation of China (No. 61402299).

## References

1. Foster, G., Isabelle, P., Plamondon, P.: Target-text Mediated Interactive Machine Translation. Machine Translation 12(1), 175–194 (1997).
2. Barrachina, S., Bender, O., Casacuberta, F., Civera, J., Cubel, E., Khadivi, S., Lagarda, A., Ney, H., Tomás, J., Vidal, E., Vilar, J. M.: Statistical Approaches to Computer-Assisted Translation. Computational Linguistics 35(1), 3–28 (2009).
3. Civera, J., Cubel, E., Lagarda, A. L., Picó, D., González, J., Vidal, E., Casacuberta, F., Vilar, J. M., Barrachina, S.: From Machine Translation to Computer Assisted Translation using Finite-State Models. In: Proceedings of EMNLP, pp. 349–356 (2004).
4. González-Rubio, J., Ortiz-Martínez, D., Benedí, J. M., Casacuberta, F.: Interactive Machine Translation using Hierarchical Translation Models. In: Proceedings of EMNLP, pp. 244–254 (2013).
5. Och, F. J., Zens, R., Ney, H.: Efficient Search for Interactive Statistical Machine Translation. In: Proceedings of EACL, pp. 287–293 (2003).
6. Tomás, J., Casacuberta, F.: Statistical Phrase-based Models for Interactive Computer-Assisted Translation. In: Proceedings of COLING/ACL, pp. 835–841 (2006).
7. Peris, Á., Domingo, M., Casacuberta, F.: Interactive Neural Machine Translation. Computer Speech and Language 45, 201–220 (2017).
8. Wuebker, J., Green, S., DeNero, J., Hasan, S., Luong, M. T.: Models and Inference for Prefix-Constrained Machine Translation. In: Proceedings of ACL, pp. 66–75 (2016).
9. Azadi, F., Khadivi, S.: Improved Search Strategy for Interactive Predictions in Computer-Assisted Translation. In: Proceedings of MT Summit, pp. 319–332 (2015).
10. Green, S., Wang, S., Chuang, J., Heer, J., Schuster, S., Manning, C. D.: Human Effort and Machine Learnability in Computer Aided Translation. In: Proceedings of EMNLP, pp. 1225–1236 (2014).
11. Koehn, P., Tsoukala, C., Saint-Amand, H.: Refinements to Interactive Translation Prediction Based on Search Graphs. In: Proceedings of ACL, pp. 574–578 (2014).
12. Ortiz-Martínez, D., García-Varea, I., Casacuberta, F.: Interactive Machine Translation based on Partial Statistical Phrase-based Alignments. In: Proceedings of RANLP, pp. 330–336 (2009).
13. Ye, N., Zhang, G., Cai, D.: Interactive-Predictive Machine Translation based on Syntactic Constraints of Prefix. In: Proceedings of COLING, pp. 1797–1806 (2016).
14. Sanchis-Trilles, G., Ortiz-Martínez, D., Civera, J.: Improving Interactive Machine Translation via Mouse Actions. In: Proceedings of EMNLP, pp. 485–494 (2008).

15. González-Rubio, J., Ortiz-Martínez, D., Casacuberta, F.: Balancing User Effort and Translation Error in Interactive Machine Translation via Confidence Measures. In: Proceedings of ACL, pp. 173–177 (2010).
16. Ueffing, N., Ney, H.: Application of Word-Level Confidence Measures in Interactive Statistical Machine Translation. In: Proceedings of EAMT, pp. 262–270 (2005).
17. González-Rubio, J., Ortiz-Martínez, D., Casacuberta F.: Active Learning for Interactive Machine Translation. In: Proceedings of EACL, pp. 245–254 (2012).
18. Ortiz-Martínez, D., García-Varea, I., Casacuberta F.: Online Learning for Interactive Statistical Machine Translation. In: Proceedings of NAACL, pp. 546–554 (2010).
19. Ortiz-Martínez, D.: Online Learning for Statistical Machine Translation. Computational Linguistics 42(1), 121–161 (2016).
20. Alabau, V., Rodríguez-Ruiz, L., Sanchis, A., Martínez-Gómez, P., Casacuberta, F.: On Multimodal Interactive Machine Translation Using Speech Recognition. In: Proceedings of the International Conference on Multimodal Interfaces, pp. 129–136 (2011).
21. Alabau, V., Sanchis, A., Casacuberta, F.: Improving On-line Handwritten Recognition in Interactive Machine Translation. Pattern Recognition 47(3), 1217–1228 (2014).
22. Alabau, V., Leiva, L. A., Ortiz-Martínez, D., Casacuberta F.: User Evaluation of Interactive Machine Translation Systems. In: Proceedings of EACL, pp. 20–23 (2012).
23. Green, S., Chuang, J., Heer, J, Manning, C. D.: Predictive Translation Memory: A Mixed-Initiative System for Human Language Translation. In: Proceedings of the Annual Association for Computing Machinery Symposium on User Interface Software and Technology, pp. 177–187 (2014).
24. Sanchis-Trilles, G., Alabau, V., Buck, C., Carl, M., Casacuberta, F., García-Martínez, M., Germann, U., González-Rubio, J., Hill, R. L., Koehn, P., Leiva, L. A., Mesa-Lao, B., Ortiz-Martínez, D., Saint-Amand, H., Tsoukala, C., Vidal, E. Interactive Translation Prediction versus Conventional Post-editing in Practice: a Study with the CasMaCat Workbench. Machine Translation 28(3), 217–235 (2014).
25. González-Rubio J., Ortiz-Martínez D., Benedí J. M., Casacuberta F.: Beyond Prefix-Based Interactive Translation Prediction. In: Proceedings of the SIGNLL Conference on Computational Natural Language Learning, pp. 198–207 (2016).
26. Federico, M., Bertoldi, N., Cettolo, M.: IRSTLM: an Open Source Toolkit for Handling Large Scale Language Models. In: Proceedings of Interspeech, pp. 1618–1621 (2008).
27. Koehn, P., Hoang, H., Birch, A., Callison-Burch, C., Federico, M., Bertoldi, N., Cowan, B., Shen, W., Moran, C., Zens, R., Dyer, C., Bojar, O., Constantin, A., Herbst, E.: Moses: Open Source Toolkit for Statistical Machine Translation. In: Proceedings of ACL, pp. 177–180 (2007).
28. Och, F. J.: Minimum Error Rate Training in Statistical Machine Translation. In: Proceedings of ACL, pp. 160–167 (2003).
29. Papineni, K., Roukos, S., Ward, T., Zhu, W. J.: BLEU: a Method for Automatic Evaluation of Machine Translation. In: Proceedings of ACL, pp. 311–318 (2002).
30. Ananthakrishnan, S., Mehay, D. N., Hewavitharana, S., Kumar, R., Roy, M., Kan, E.: Lightly Supervised Word-Sense Translation-Error Detection and Resolution in an Interactive Conversational Spoken Language Translation System. Machine Translation 29(1), 25–47 (2015).
31. Huang, G., Zhang, J., Zhou, Y., Zong, C.: A New Input Method for Human Translators: Integrating Machine Translation. In: Proceedings of IJCAI, pp. 1163–1169 (2015).
32. Cheng, S., Huang, S., Chen, H., Dai, X., Chen, J.: PRIMT: A Pick-Revise Framework for Interactive Machine Translation. In: Proceedings of NAACL, pp. 1240–1249 (2016).