

Interactive-Predictive Translation based on Multiple Word-Segments

Miguel DOMINGO, Álvaro PERIS, Francisco CASACUBERTA

Pattern Recognition and Human Language Technology Research Center
Camino de Vera s/n, 46022 Valencia, Spain

midobal@prhlt.upv.es, lvapeab@prhlt.upv.es, fcn@prhlt.upv.es

Abstract. Current machine translation systems require human revision to produce high-quality translations. This is achieved through a post-editing process or by means of an interactive human–computer collaboration. Most protocols belonging to the last scenario follow a left-to-right strategy, where the prefix of the translation is iteratively increased by successive validations and corrections made by the user. In this work, we propose a new interactive protocol which allows the user to validate all correct word sequences in the translation generated by the system, breaking the left-to-right barrier. We evaluated our proposal through simulated experiments, obtaining large reductions of the human effort.

Keywords: machine translation, computer-assisted translation, interactive-predictive machine translation

1 Introduction

Machine Translation (MT) technology is still far from producing perfect translations (Dale, 2016). Therefore, translation errors must be corrected by a human in a later post-editing stage.

The Interactive-Predictive Machine Translation (IMT) field arose as an alternative to classic post-editing systems, aiming to reduce human post-editing effort and increase efficiency. This paradigm strives for combining the knowledge of a human translator and the efficiency of an MT system. Notable contributions to IMT technology were carried out around the *TransType* (Foster et al., 1997; Langlais and Lapalme, 2002), *TransType2* (Barrachina et al., 2009; Casacuberta et al., 2009); and *CasMaCat* (Martínez-Gómez et al., 2012; Alabau et al., 2013; González-Rubio et al., 2013; Sanchis-Trilles et al., 2014) projects, among others (Koehn, 2009; Huang et al., 2012; Cai et al., 2013; Green et al., 2014; Torregrosa et al., 2014; Azadi and Khadivi, 2015; Marie and Max, 2015).

Especially interesting is the so-called prefix-based IMT (Barrachina et al., 2009). In this approach, the user corrected the first wrong word (from left-to-right) of the

translation suggested by the system. Then, the system proposed an alternative hypothesis, compatible with the user feedback. A cumbersome phenomenon noticed in this protocol happened when the non-validated part of the sentence contained correct words. If such words were modified by the system in following predictions, the user had to edit words that were correct in previous iterations. Therefore, the effort made by the user was increased and the system had an annoying behavior.

To overcome this weakness, we propose a new IMT approach which allows the user to select, at each interaction, all correctly translated word segments. Hence, correct parts of the current translation are kept in successive hypothesis produced during the human-machine interaction, reducing the number of corrections required and avoiding the aforementioned issue. This approach relies on the idea from González-Rubio et al. (2016) of breaking down the prefix constraint.

The proposed protocol shares some similarities with Marie and Max (2015) in that we select word segments from a translation hypothesis. However, on the one hand, our protocol contains more types of user interactions such as word corrections and word deletions (see Section 2); and, on the other hand, we have different goals in mind: Marie and Max (2015) aim at increasing translation quality with the help of a human user, and we aim at reducing the human effort of generating a translation in an IMT framework.

The rest of this paper is structured as follows: Section 2 describes our segment-based IMT approach. After that, in Section 3, we report the experiments conducted in order to assess our proposal and the results of those experiments. Finally, conclusions of the work are drawn in Section 4.

2 Segment-Based Search

The goal of the IMT protocol developed in this work is to offer more freedom to the human agent, empowering the selection of the correct segments of a translation hypothesis. To achieve this, we allow the user to select, remove, or replace parts of a translation suggestion. The system then reacts to this human feedback, producing a new compatible hypothesis. Fig. 1 shows an example of an IMT session using the proposed segment-based approach.

2.1 Statistical Framework

Barrachina et al. (2009) proposed an statistical framework for the prefix-based IMT approach, where human and computer iteratively collaborated for translating a source sentence \mathbf{x} . In this framework, at the beginning of the process, the system proposes a translation hypothesis \mathbf{y} . Then, the user searches, from left-to-right, the first wrong word in \mathbf{y} and corrects it. With this action, the user defines a valid translation prefix $\hat{\mathbf{p}}$. At the next iteration, the system generates a suffix $\tilde{\mathbf{s}}$ that completes $\hat{\mathbf{p}}$ in order to (hopefully) obtain a better translation of \mathbf{x} : $\mathbf{y}' = \hat{\mathbf{p}}\tilde{\mathbf{s}}$. This process is repeated until the user accepts the complete suggestion of the system. At each iteration, $\tilde{\mathbf{s}}$ is obtained as the most probable of all possible suffixes \mathbf{s} , given the prefix $\hat{\mathbf{p}}$ and the source sentence \mathbf{x} :

$$\tilde{\mathbf{s}} = \arg \max_{\mathbf{s}} Pr(\mathbf{s} | \mathbf{x}, \hat{\mathbf{p}}) \quad (1)$$

source (x): Et la question n ' a pas encore été évaluée chez les patients atteints de cancer gastrique

target translation (\hat{y}): And the issue has not been evaluated in gastric cancer patients

IT-0	T	And the issue has not yet been investigated among patients with gastric cancer
IT-1	U	[And the issue has not been] evaluated among patients with [gastric cancer]
	T	[And the issue has not been evaluated] not in [gastric cancer] patients with
IT-2	U	[And the issue has not been evaluated in gastric cancer patients] #
	T	[And the issue has not been evaluated in gastric cancer patients]
END	U	And the issue has not been evaluated in gastric cancer patients

Fig. 1: Segment-based IMT session to translate a French sentence into English. At the initial iteration (*IT-0*), the system suggests an initial translation. Then, at iteration 1, the user selects those segments to keep (“*And the issue has not*”, “*been*” and “*gastric cancer*”); deletes a word (“*yet*”); and substitutes “*investigated*” by “*evaluated*”, which is added to the segment. With this information, the system suggests a new hypothesis. Similarly, at iteration 2, the user selects new valid segments (“*in*” and “*patients*”), deletes words that are in the middle of two segments (“*not*”), and inputs an *end of sentence* mark (illustrated as “*#*”). The session ends when the user accepts the last translation suggested by the system.

This equation can be straightforwardly rewritten as:

$$\tilde{s} = \arg \max_{\mathbf{s}} Pr(\hat{\mathbf{p}}, \mathbf{s} | \mathbf{x}) \quad (2)$$

Therefore, at each iteration, the process consists of a regular search in the space of the translations but constrained by the prefix $\hat{\mathbf{p}}$.

The protocol proposed in our work follows this iterative procedure but, at each iteration, the user is free to validate all correct subsequences of words (segments) from \mathbf{y} . The user has also the possibility of deleting all words located between two segments (merging both segments into one), and either correcting a wrong word (as in the prefix-based approach) or inserting a new word between two segments.

Let $\hat{\mathbf{f}} = \hat{\mathbf{f}}_1, \dots, \hat{\mathbf{f}}_N$ be a feedback signal, where $\hat{\mathbf{f}}_1, \dots, \hat{\mathbf{f}}_N$ is the sequence of N segments validated by the user in an interaction (including a one-word segment with the new word). The goal is to generate a sequence $\mathbf{h} = \tilde{\mathbf{h}}_1, \dots, \tilde{\mathbf{h}}_N$ of new translation segments (an $\tilde{\mathbf{h}}_i$ for each pair of validated segments $\hat{\mathbf{f}}_i, \hat{\mathbf{f}}_{i+1}$; being $1 \leq i < N$) to obtain a (hopefully) better translation of \mathbf{x} : $\mathbf{y}' = \hat{\mathbf{f}}_1, \tilde{\mathbf{h}}_1, \dots, \hat{\mathbf{f}}_N, \tilde{\mathbf{h}}_N$. In our statistical framework, the best translation segments are obtained as:

$$\tilde{\mathbf{h}}_1, \dots, \tilde{\mathbf{h}}_N = \arg \max_{\mathbf{h}_1, \dots, \mathbf{h}_N} Pr(\mathbf{h}_1, \dots, \mathbf{h}_N | \mathbf{x}, \hat{\mathbf{f}}_1, \dots, \hat{\mathbf{f}}_N) \quad (3)$$

which can be rewritten as:

$$\tilde{\mathbf{h}}_1, \dots, \tilde{\mathbf{h}}_N = \arg \max_{\mathbf{h}_1, \dots, \mathbf{h}_N} Pr(\hat{\mathbf{f}}_1, \mathbf{h}_1, \dots, \hat{\mathbf{f}}_N, \mathbf{h}_N | \mathbf{x}) \quad (4)$$

This last equation is very similar to the classical prefix-based IMT equation (Eq. (1)), with the main difference being that the search process in Eq. (1) is limited to the space of suffixes constrained by $\hat{\mathbf{p}}$, while the search in Eq. (4) is in the space of possible substrings of the translations of \mathbf{x} , constrained by the sequence of segments $\hat{\mathbf{f}}_1, \dots, \hat{\mathbf{f}}_N$.

3 Experiments

3.1 Corpora

We tested our proposal in four tasks from different domains: the **EMEA** corpus¹ (Tiedemann, 2009), formed by documents from the *European Medical Agency*; the **EU** corpus (Barrachina et al., 2009), extracted from the *Bulletin of the European Union*; the **TED** corpus² (Federico et al., 2011), a collection of recordings of public speeches covering a variety of topics; and the **Xerox** corpus (Barrachina et al., 2009), extracted from *Xerox* printer manuals. To the best of our knowledge, excluding EMEA, all corpora have been used in previous IMT works (Tomás and Casacuberta, 2006; Barrachina et al., 2009; González-Rubio et al., 2013). The partition sets used in this work are the same than those used in the aforementioned works.

All datasets have been tokenized by means of the standard tool provided with the *Moses* toolkit (Koehn et al., 2007)—exempting Chinese sentences, which were split into words using the Stanford word segmenter (Tseng et al., 2005). Sentences have been kept truecased, except for the Zh–En language pair, since Chinese has no case information. Table 1 shows the corpora main features.

Table 1: Corpora statistics. K denotes *thousands* and M *millions*. $|S|$ stands for *number of sentences*, $|W|$ for *number of words* and $|V|$ for *size of the vocabulary*.

		EMEA (Fr/En)	EU (Es/En)	TED (Zh/En)	Xerox (Es/En)
Train	$ S $	1.1M	214K	106.9K	55.6K
	$ W $	14.3M/17.0M	6M/5.4M	1.9M/2.1M	750K/665K
	$ V $	71K/80K	84K/70K	55K/41.7K	16.8K/14K
Dev.	$ S $	500	400	934	1012
	$ W $	12K/10K	12K/10K	21.5K/20.1K	16K/14.4K
	$ V $	2.9K/2.7K	3K/2.7K	3.8K/3.2K	1.8K/1.6K
Test	$ S $	1K	800	1.6K	1.1K
	$ W $	27K/21K	23K/20K	33.2K/31.9K	10.1K/8.4K
	$ V $	4.5K/4.5K	4.7K/4.2K	4.5K/3.7K	2K/1.9K

3.2 Metrics

The quality of our interactive protocol is assessed according to the following metrics:

Word Stroke Ratio (WSR) (Tomás and Casacuberta, 2006): Measures the number of words edited by the user, normalized by the number of words in the final translation. In this work, we assume that the edition of a word is considered to have a constant cost (one word stroke) independently of its length.

¹ <http://www.statmt.org/wmt14/medical-task/>

² <https://wit3.fbk.eu/mt.php?release=2012-03-test>

Mouse Action Ratio (MAR) (Barrachina et al., 2009): Measures the number of mouse actions made by the user, normalized by the number of characters in the final translation. In classic IMT, the user makes a mouse action each time she needs to edit a word (to position the prompt), and one more per sentence to validate the translation. In the protocol proposed in this work, in addition to those mouse actions, the user makes two actions each time she validates a segment (clicking at the beginning and at the end of the segment), and two more each time she deletes some words located between segments³ (same procedure as selecting segments but using the right button of the mouse).

Conceptually, WSR accounts for the physical effort of typing corrections, while MAR accounts for the cognitive effort of the supervision process (Macklovitch et al., 2005).

Additionally, to evaluate the quality of the initial translations, we have used the following well-known metric:

BiLingual Evaluation Understudy (BLEU) (Papineni et al., 2002): computes the geometric average of the modified n -gram precision, multiplied by a factor that penalizes short sentences.

3.3 Implementation

Our implementation of the segment-based IMT protocol is based on the `Moses` toolkit (Koehn et al., 2007). We profit from the feature that allows to bring external knowledge to the decoder by means of an *XML Markup* language (see Fig. 2 for an example), for validating the translation of parts of a sentence without changing the models. The decoder has an XML markup scheme that allows us to plug in the translation of parts of a sentence without changing the models. More precisely, we use the *exclusive* mode, which only takes into account the given translation for a part of a sentence—ignoring any phrases from the phrase table that overlaps with that span. With this, we can constrain the search process to follow Eq. (4).

```
<x translation = "And the issue has not been evaluated"> Et la question n ' a pas encore été
évaluée </x><wall/> chez les patients atteints de <x translation = "gastric cancer"> cancer
gastrique </x><wall/>
```

Fig. 2: Example of a sentence in XML markup language (corresponding to the sentence of the first iteration of Fig. 1), specifying the desired translation for some parts of the sentence: *Et la question n ' a pas encore été évaluée* must be translated as *And the issue has not been evaluated*, and *cancer gastrique* must be translated as *gastric cancer*. The tag `<wall/>` indicates to the decoder that those segments should not be reordered.

We implemented a prototype that manages the interaction between a human agent and the MT system. This is an iterative process in which the prototype, by means of the

³ One mouse action is enough for selecting or deleting a one-word segment (in which case, the user would simply click on the word).

XML markup language, takes into account the feedback provided by the user, obtains a translation with `Moses`, and suggests the new hypothesis. All this takes place at the end of each iteration, with an average response time of 90 ms⁴ per iteration. According to Nielsen (1993), this time is below “*the limit for having the user feel that the system is reacting instantaneously*”.

At each one of these iterations, the user has three different ways of interacting with the system (see Section 2). Such interactions affect differently in the generation of the new XML markup sentence:

Segment selection: for each segment selected by the user, we align the words of that segment with their correspondent source words (phrase alignments), and generate an XML tag to plug in that segment (the desired translation) to those source words.

Word deletion: in the same fashion as with segments, for each word to delete, we align that word with its correspondent source words and generate a new XML tag, indicating that we want to obtain an empty translation.

Word correction: each time the user corrects a word or inserts a new one, we align the new word with its correspondent source words using a hidden markov alignment model (Vogel et al., 1996).

All the MT systems used in this work were trained with the standard configuration of `Moses`, with the weights of the log-linear model being optimized by means of the Minimum Error Rate Training (MERT) procedure (Och, 2003). Lastly, a 5-gram word-based language model was estimated on the target side of the parallel corpora, using the improved KneserNey smoothing (Chen and Goodman, 1996), by means of the SRILM toolkit (Stolcke, 2002).

For the implementation of the classic prefix-based IMT systems, we made the word graph exploration and the best suffix generation for a given prefix following the procedure described by Barrachina et al. (2009): We generated a word graph for each sentence to translate. After that, treating the word graph as a weighted finite-state automaton, we parsed the prefix over it, from the initial state to any other intermediate state, to find the best path that accounts for the prefix. Finally, we obtained the corresponding translation for the best path from the intermediate state to the finale state. Therefore, our implementation of prefix-based IMT is consistent with Barrachina et al. (2009), considering that we generate word graphs with the current SMT state-of-the-art `Moses` toolkit.

3.4 Evaluation on a Simulated Environment

Since the evaluation with human agents is too slow and expensive to be applied frequently during system development, we carried out an automatic evaluation with simulated users. For this evaluation, we considered the references in the corpora as the translations the user desires. Furthermore, without loss of generality and for the sake of simplicity, we assumed that the user always corrected the left-most wrong word.

At each iteration of the IMT session, we selected those segments that were common with the reference. After that, following a left-to-right order, we compared each word of

⁴ Tested on a machine with an Intel i5 CPU at 3.1 GHz.

the current translation with those of the reference. When we found a different word in translation and reference, if that reference word was the first one of the next selected segment, we deleted all the words between those two segments; otherwise, we input that word (merging all previous segments into one). Once translation and reference were the same, we moved on to the next sentence.

3.5 Results

Table 2 shows the user-effort results of our segment-based protocol against the prefix-based approach. Prefix-based results were obtained following the work of Barrachina et al. (2009) and are similar to those reported on the literature (Tomás and Casacuberta, 2006; Barrachina et al., 2009; González-Rubio et al., 2013). The quality of the initial translation is also displayed as an indicative of the difficulty of each task. Our proposal clearly improves prefix-based IMT in terms of user physical effort of typing corrections. The WSR is always reduced, yielding diminishes up to 29 points.

Table 2: Results of our segment-based IMT proposal, in comparison with the prefix-based approach. The quality of the initial translation is shown as an indicative of the difficulty of each task. All values are reported as percentages.

Corpus	Language	BLEU	Prefix-Based		Segment-Based	
			WSR	MAR	WSR	MAR
EMEA	Fr-En	31.3	57.8	12.4	34.4	18.8
	En-Fr	30.2	58.4	12.5	40.4	16.3
EU	Es-En	48.2	45.6	10.2	28.3	15.0
	En-Es	48.7	44.6	9.7	29.8	13.5
TED	Zh-En	11.7	83.1	22.4	54.1	28.3
	En-Zh	8.7	86.3	55.7	59.2	72.4
Xerox	Es-En	54.5	35.8	10.5	23.2	16.9
	En-Es	62.2	28.3	7.9	22.1	12.5

This reduction of typing effort comes with an increase in the number of mouse actions (from 4 up to 6.5 points of MAR), which is always smaller than the effort reduction. An exception to this comes with the En-Zh language pair since, due to Chinese nature, words have fewer number of characters, which penalizes MAR metric. This penalization results in a greater increase in MAR, although this increase is still smaller than the effort reduction. Moreover, as mentioned before, WSR and MAR account for different phenomena and thus have different cost from a human point of view (Macklovitch et al., 2005). Therefore, the physical effort is substantially decreased, while the cognitive one is slightly increased. Nonetheless, we need to test these considerations with real human users before reaching to categorical conclusions.

4 Conclusions

In this work, we have proposed a new IMT approach that overcomes the classic prefix-based IMT limitation of only correcting the prefix. Our proposal allows the user to select all correct word segments each time the system proposes a new translation. The system leverages this additional knowledge for offering more enlightened hypothesis. Hence, the human typing effort should be reduced.

We tested the proposal in a simulated environment, which confirmed that our approach effectively reduces the physical effort required, at the expense of a slight increase in the cognitive effort. As future work, we should test the improvements of our proposal with real users in order to obtain actual measures of the effort reduction.

Acknowledgments

The research leading to these results has received funding from the Ministerio de Economía y Sostenibilidad (MINECO) under project SmartWays (grant agreement RTC-2014-1466-4), and Generalitat Valenciana under project ALMAMATER (grant agreement PROMETEOII/2014/030).

References

- Alabau, Vicent, Ragnar Bonk, Christian Buck, Michael Carl, Francisco Casacuberta, Mercedes García-Martínez, Jesús González-Rubio, Philipp Koehn, Luis A. Leiva, Bartolomé Mesa-Lao, Daniel Ortiz-Martínez, Hervé Saint-Amand, Germán Sanchis-Trilles, and Chara Tsoukala (2013). “CASMACAT: An Open Source Workbench for Advanced Computer Aided Translation”. In: *The Prague Bulletin of Mathematical Linguistics* 100, pp. 101–112.
- Azadi, Fatemeh and Shahram Khadivi (2015). “Improved Search Strategy for Interactive Machine Translation in Computer-Assisted Translation”. In: *Proceedings of Machine Translation Summit XV*, pp. 319–332.
- Barrachina, Sergio, Oliver Bender, Francisco Casacuberta, Jorge Civera, Elsa Cubel, Shahram Khadivi, Antonio Lagarda, Hermann Ney, Jesús Tomás, Enrique Vidal, and Juan-Miguel Vilar (2009). “Statistical Approaches to Computer-Assisted Translation”. In: *Computational Linguistics* 35, pp. 3–28.
- Cai, Dongfeng, Hua Zhang, and Na Ye (2013). “Improvements in Statistical Phrase-Based Interactive Machine Translation”. In: *Proceedings of the International Conference on Asian Language Processing*, pp. 91–94.
- Casacuberta, Francisco, Jorge Civera, Elsa Cubel, Antonio L. Lagarda, Guy Lapalme, Elliott Macklovitch, and Enrique Vidal (2009). “Human Interaction for High-quality Machine Translation”. In: *Communications of the Association for Computing Machinery* 52.10, pp. 135–138.
- Chen, Stanley F. and Joshua Goodman (1996). “An Empirical Study of Smoothing Techniques for Language Modeling”. In: *Proceedings of the Annual Meeting on Association for Computational Linguistics*, pp. 310–318.

- Dale, Robert (2016). “How to make money in the translation business”. In: *Natural Language Engineering* 22.2, pp. 321–325.
- Federico, Marcello, Luisa Bentivogli, Michael Paul, and Sebastian Stüker (2011). “Overview of the IWSLT 2011 evaluation campaign”. In: *International Workshop on Spoken Language Translation*, pp. 11–27.
- Foster, George, Pierre Isabelle, and Pierre Plamondon (1997). “Target-Text Mediated Interactive Machine Translation”. In: *Machine Translation* 12, pp. 175–194.
- González-Rubio, Jesús, Daniel Ortiz-Martínez, José-Miguel Benedí, and Francisco Casacuberta (2013). “Interactive Machine Translation using Hierarchical Translation Models”. In: *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pp. 244–254.
- González-Rubio, Jesús, José-Miguel Benedí, and Francisco Casacuberta (2016). “Beyond Prefix-Based Interactive Translation Prediction”. Unpublished results.
- Green, Spence, Jason Chuang, Jeffrey Heer, and Christopher D. Manning (2014). “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.
- Huang, Chung-chi, Ping-che Yang, Keh-jian Chen, and Jason S. Chang (2012). “TransAhead: A Computer-Assisted Translation and Writing Tool”. In: *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics*, pp. 352–356.
- Koehn, Philipp (2009). “A Web-Based Interactive Computer Aided Translation Tool”. In: *Proceedings of the International Joint Conference on Natural Language Processing*, pp. 17–20.
- Koehn, Philipp, Hieu Hoang, Alexandra Birch, Chris Callison-Burch, Marcello Federico, Nicola Bertoldi, Brooke Cowan, Wade Shen, Christine Moran, Richard Zens, Chris Dyer, Ondřej Bojar, Alexandra Constantin, and Evan Herbst (2007). “Moses: Open Source Toolkit for Statistical Machine Translation”. In: *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, pp. 177–180.
- Langlais, Philippe and Guy Lapalme (2002). “TransType: Development-Evaluation Cycles to Boost Translator’s Productivity”. In: *Machine Translation* 17.2, pp. 77–98.
- Macklovitch, Elliot, Nam-Trung Nguyen, and Roberto Silva (2005). *User evaluation report*. Tech. rep. Transtype2 (ISR-2001-32091).
- Marie, Benjamin and Aurélien Max (2015). “Touch-Based Pre-Post-Editing of Machine Translation Output”. In: *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pp. 1040–1045.
- Martínez-Gómez, Pascual, Germán Sanchis-Trilles, and Francisco Casacuberta (2012). “Online Adaptation Strategies for Statistical Machine Translation in Post-Editing Scenarios”. In: *Pattern Recognition* 45.9, pp. 3193–3203.
- Nielsen, Jakob (1993). *Usability Engineering*. Morgan Kaufmann Publishers Inc. ISBN: 0125184050.
- Och, Franz Josef (2003). “Minimum Error Rate Training in Statistical Machine Translation”. In: *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, pp. 160–167.

- Papineni, Kishore, Salim Roukos, Todd Ward, and Wei-Jing Zhu (2002). “BLEU: a Method for Automatic Evaluation of Machine Translation”. In: *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, pp. 311–318.
- Sanchis-Trilles, Germán, Vicent Alabau, Christian Buck, Michael Carl, Francisco Casacuberta, Mercedes García-Martínez, Ulrich Germann, Jesús González-Rubio, Robin Hill, Philipp Koehn, Luis A. Leiva, Bartolomé Mesa-Lao, Daniel Ortiz-Martínez, Hervé Saint-Amand, Chara Tsoukala, and Enrique Vidal (2014). “Interactive Translation Prediction vs. Conventional Post-editing in Practice: A Study with the CasMaCat Workbench”. In: *Machine Translation 28.3–4*, pp. 217–235.
- Stolcke, Andreas (2002). “SRILM - An extensible language modeling toolkit”. In: *Proceedings of the International Conference on Spoken Language Processing*, pp. 257–286.
- Tiedemann, Jörg (2009). “News from OPUS - A Collection of Multilingual Parallel Corpora with Tools and Interfaces”. In: *Recent Advances in Natural Language Processing*. Vol. V, pp. 237–248.
- Tomás, Jesús and Francisco Casacuberta (2006). “Statistical Phrase-Based Models for Interactive Computer-Assisted Translation”. In: *Proceedings of the International Conference on Computational Linguistics/Association for Computational Linguistics*, pp. 835–841.
- Torregrosa, Daniel, Mikel L. Forcada, and Juan Antonio Pérez-Ortiz (2014). “An Open-Source Web-Based Tool for Resource-Agnostic Interactive Translation Prediction”. In: *Prague Bulletin of Mathematical Linguistics* 102, pp. 69–80.
- Tseng, Huihsin, Pichuan Chang, Galen Andrew, Daniel Jurafsky, and Christopher Manning (2005). “A Conditional Random Field Word Segmenter”. In: *Proceedings of the Special Interest Group of the Association for Computational Linguistics Workshop on Chinese Language Processing*, pp. 168–171.
- Vogel, Stephan, Hermann Ney, and Christoph Tillmann (1996). “HMM-based Word Alignment in Statistical Translation”. In: *Proceedings of the Conference on Computational Linguistics*. Vol. 2, pp. 836–841.

Received April 29, 2016 , accepted May 15, 2016