

An SMT-driven Authoring Tool

*Sriram Venkatapathy*¹ *Shachar Mirkin*¹

(1) Xerox Research Centre Europe

`sriram.venkatapathy@xrce.xerox.com, shachar.mirkin@xrce.xerox.com`

ABSTRACT

This paper presents a tool for assisting users in composing texts in a language they do not know. While Machine Translation (MT) is pretty useful for understanding texts in an unfamiliar language, current MT technology has yet to reach the stage where it can be used reliably without a post-editing step. This work attempts to make a step towards achieving this goal. We propose a tool that provides suggestions for the continuation of the text in the source language (that the user knows), creating texts that can be translated to the target language (that the user does not know). In terms of functionality, our tool resembles text prediction applications. However, the target language, through a Statistical Machine Translation (SMT) model, drives the composition and not only the source language. We present the user interface and describe the considerations that underline the suggestion process. A simulation of user interaction shows that composition speed can be substantially reduced and provides initial positive feedback as to the ability to generate better translations.

KEYWORDS: Statistical Machine Translation.

1 Introduction

A common task in today's multilingual environments is to compose texts in a language that the author is not fluent in or not familiar with at all. This could be a prospective tourist sending an email to a hotel abroad or an employee of a multinational company replying to customers who speak another language than his own. A solution to such problems is to enable handling multilingual text-composition using Machine Translation technology. The user composes a text in his own (*source*) language and it is then automatically translated to the *target* language. Machine translation technology can be used rather effectively to interpret texts in an unfamiliar language. Automatic translations are often good enough for understanding the general meaning of the text, even if they do not constitute a perfect translation or are not perfectly written. On the other hand, the state-of-the-art MT technology is often not suitable for directly composing foreign-language texts. Erroneous or non-fluent texts may not be well-received by the readers. This is especially important in business environments, where the reputation of the company may be harmed if low-quality texts are used by it. The output of the MT system must therefore go through a *post-editing* process before the text can be used externally. That is, a person who is fluent in both the source and the target languages must review and correct the target text before it is sent or displayed. Obviously, such a post-editing step, which requires knowledge of both languages, makes the process slow, expensive and even irrelevant in many cases.

In this work we present an approach driven by Statistical Machine Translation for enabling users to compose texts that will be translated more accurately to a language unfamiliar to them. This is a step towards composition of texts in a foreign language that does not depend on knowledge of that language.

The composition of the text in the desired language is a two-step process:

1. Through an interactive interface the user is guided to **quickly** compose a text in his native language that *can be* more **accurately** translated to the target language. The composition guidance is powered by the SMT-system.
2. The text is translated by the SMT system.

Hence, the SMT system plays a dual role in this method, for interactive composition of source-language texts and for their translation into the target language. Its role in the composition of the source by the user is what enables improving the accuracy of the translation. The interface prompts the user to choose those phrases that can be translated more accurately by the SMT system. In contrast, a text composed directly by the user might contain terminology or sentence structure that the SMT system cannot successfully translate.

Apart from improving the accuracy of the translation, this method may also improve composition speed. This is made possible by the interface that allows the user to click on desired phrases to extend the current text, thereby saving typing. Composition speed is further improved by displaying complete sentences from the translation memory that are similar to the user's input at any given time and which the user can simply select to be translated.

2 Related work

To our knowledge, there has been no previous work where interactive authoring is driven by an SMT model.

Interactive tools for MT-targeted authoring were proposed in several works. (Dymetman et al., 2000) suggested a method for assisting monolingual writers in the production of multilingual documents based on a parallel grammar; (Carbonell et al., 2000) propose a tool to produce controlled texts in the source language; (Choumane et al., 2005) provide interactive assistance to authors based on manually defined rules, for tagging the source text in order to reduce its syntactic ambiguity. In contrast, our method does not depend on predefined templates or rules for a constrained language, but is tightly coupled with the SMT models which are learnt automatically from parallel corpora.

Tools for computer-assisted translation (CAT) are meant to increase the productivity of translators. A fundamental difference between these and our suggested tool is that CAT systems are designed for translators and operate on the **target** side of the text, assuming their users know both the source and target languages, or at least the target language. For example, (Koehn and Haddow, 2009) proposed a CAT system ¹ which suggests words and phrases for target-side sentence completion based on the phrase table. (Koehn, 2010) and (Hu et al., 2011) propose translation by monolinguals, but also rely on users fluent in the target language.

Our tool bears resemblance to *text prediction* applications, such as smart-phone keyboards. In comparison to these, our tool suggests the next phrase(s) based both on the user's input and the translatability of these phrases, according to an SMT model. This naturally stems from the different purpose of our tool - to compose a text in **another** language.

3 Interactive interface

In this section, we present the interactive interface that enables users to compose texts in a language they do not know (see Figure 1). To compose a text, the user starts typing in his native language (English in this example) in the top text-box. The interface allows the user to perform the following primary operations at any point of time: (1) **Phrase selection**: Select one of the phrase suggestions to expand the sentence. The number of suggested phrases is controlled by the user through the interface, (2) **Sentence selection**: Select an entire sentence from a list of sentences similar to the partial input. Words that match the typed text are highlighted, and (3) **Word composition**: The user may go on typing if he does not want to use any of the provided suggestions. Any selected text is also editable.

Whenever the space button is hit, new phrase and sentence suggestions are instantly shown. A mouse click on a suggested sentence copies it to the top text box, replacing the already typed text; a selection of a phrase appends it to the typed text. Phrases are ordered by their estimated "appropriateness" to the typed text (see details in Section 4). Throughout the process, the partial translation (to French in our example) can be shown with the toggle show/hide translation button.

4 SMT-Driven authoring

The goal of our work is to enable users to compose a text in a language they know that can be translated accurately to a language which they do not know. We assume that when starting to type, the user typically has some intended message in mind, whose meaning is set, but not its exact wording. This enables us to suggest text continuations that will preserve the intended meaning, but phrased in a way that it will be better translated. This is achieved by a guidance method that is determined by three factors:

¹www.caitra.org

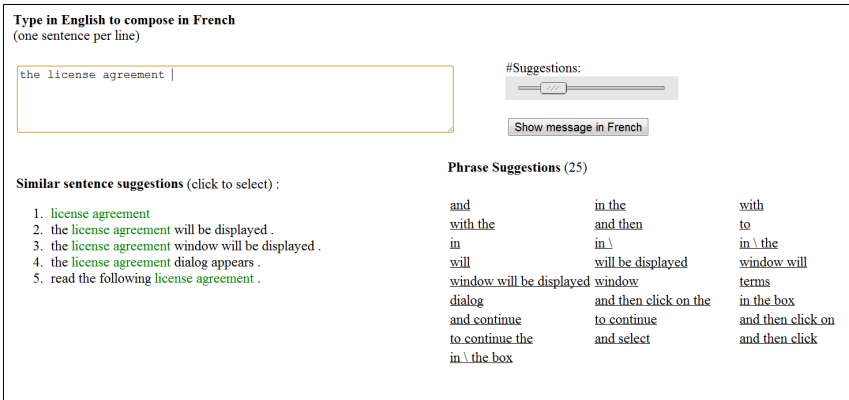


Figure 1: The user interface of the authoring tool.

1. **Fluency:** The phrase suggestions shown should be fluent with respect to the input already provided by the user. As our setting is interactive, the fluency of the source text must be maintained at any given time.
2. **Translatability:** The ability of the SMT system to translate a given source phrase. This is a factor controlled by the authoring tool by proposing phrases that can be translated more accurately, thereby moving towards a better translation of the intended text.
3. **Semantic distance:** The semantic distance between the composed source text and the suggestions. This criterion is required in order to ensure our suggestions are not simply those that are easy to translate, thus preventing a deviation from the meaning of the intended text (in contrast to deviating from the wording). A high distance corresponds to a different meaning of the composed message relative to the intended one. In such cases, the SMT system cannot be expected to generate a translation whose meaning is close to the desired one.

To rank proposed phrases we use a metric whose aim is to both minimize the semantic deviation of suggested phrases from the already typed text and to maximize the translatability of the text. That, while maintaining the fluency of the sentence. This metric is a weighted product of individual metrics for each of the three above factors. Briefly, to maintain source fluency we suggest only those phrases whose prefix overlaps with the suffix of the user's partial input. We use the SRILM toolkit (Stolcke, 2002) to obtain the per-word-perplexity of each suggested phrase, and normalize it by the maximal perplexity of the language model. This yields a normalized score over different phrase lengths. We assessed two metrics for estimating phrase translatability. The first is based on conditional entropy (*CE*), following (DeNero et al., 2006) and (Moore and Quirk, 2007). The idea is that a source phrase is more translatable if it occurs frequently in the training data and has more valid options for translation. The second is the maximum translation score (*MTS*), computed from the translation features in the SMT phrase table, maximized over all phrase-table entries for a given source phrase. To minimize meaning deviation, we compute the averaged DICE coefficient (Dice, 1945) between the suggested phrase and the already-typed input, which measures the tendency of their words to co-occur

according to corpus statistics. We applied a sigmoid function to the translatability scores to bring them to [0-1] range, and used the square root of the DICE score in order to scale it to a more similar range of the other scores.

5 Implementation

The interactive nature of our proposal requires instant response from the authoring tool. This is a critical requirement, even if large translation models and corpora are employed, as users would find it useless if response to their actions is not instantaneous. To enable immediate suggestions of phrases, we create an inverted index of the phrase table, indexing all prefixes of the source phrases in the table. This enables instantly providing suggestion by retrieving from the index all phrase table entries which have the typed few words as prefix. Indexing and retrieval in our implementation are done using the Lucene search engine².

6 Evaluation

We empirically measured the utility of the interface for the task of composing texts in an unfamiliar language by computing the cost of the text composition and the accuracy of the translation of the composed text. Ideally, this needs to be done manually by human evaluators, yet at this stage we performed the evaluation through a simulator that emulates a human who is using the interface. The simulator's goal is to compose a text using the authoring tool while remaining semantically close to the intended text the user had in mind. For our purposes, the intended texts are existing corpus sentences which we try to reconstruct. The simulator attempts to reconstruct these texts by making the least possible 'effort'. If it is possible, an entire sentence is selected; otherwise the longest possible matching phrase is chosen. If no such match is found, words from the intended text are copied to the sentence being recomposed, which is equivalent to a composition by the user.

We applied the simulation of two datasets for our experiments: (1) Technical Manuals Dataset (TM), and (3) News commentary Dataset - WMT2007 (NC).

Evaluating composition cost To assess composition cost we used a simulator that tries to reconstruct the intended text exactly. That is, the simulator selects a phrase suggestion only if it is identical to the following word(s) of the intended text. Let us assume, for instance, that the intended text is *'the license agreement will be displayed next'* and the text *'the license agreement'* has already been composed. The simulator can either select a sentence identical to the entire intended text or to select the longest prefix of the phrase *'will be displayed next'* from among the suggested phrases. The results when applying this simulator are presented in Figure 2. As shown in the figure, word composition is reduced when the tool is being used, and unsurprisingly, more suggestions yield a greater save in composition cost. We further see that the *CE* metric is preferable over *MTS*, better ranking the suggestions which leads to their selection rather their composition.

Evaluating translation accuracy By design, the simulator mentioned above cannot assess potential gains in translation accuracy. To achieve that we must allow it to compose texts that are different from the original ones. For that purpose we created additional simulators, which can – to some extent – modify the intended text, generating simple paraphrases of it by allowing substitution of function words or content words by their synonyms. For instance,

²<http://lucene.apache.org>

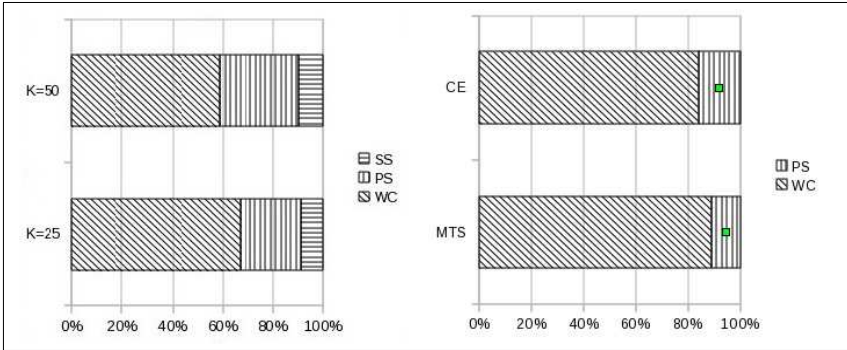


Figure 2: Percentage of word compositions (WC), phrase selections (PS) and Sentence Selection (SS) using different values of k on a technical manuals dataset (left) and using different translatability metrics (CE and MTS) on the News Commentary dataset (right).

the words ‘*will be displayed next*’ in the intended text may be replaced in this simulation by the phrase suggestion ‘*is shown next*’. We applied these simulators on the test sets, but very few replacements occurred. This result does not allow us to report translation performance scores at this stage and calls for rethinking the simulation methodology and for manual evaluation. Yet, in some cases changes were made by the simulator and resulted with improved translations, as measured by the BLEU score (Papineni et al., 2002). As an example, in the sentence ‘*The candidate countries want to experience quick economic growth ...*’, the word *countries* was replaced by *states*, resulting in a better BLEU score (0.2776 vs. 0.2176).

7 Conclusions and future work

This work represents a step towards more accurate authoring of texts in a foreign language. While not constituting at this stage an alternative to post-editing, it may enable reducing the effort incurred by such a process. We proposed a tool for assisting the user during the composition of a text in his native language where next-words suggestions are driven by an SMT-model, such that the eventual translation would be better than if the user had written the text by himself. Through our evaluation we have demonstrated that composition speed can be increased and exemplified a better translation that is produced when using the tool.

Thus far, our approach and tool were evaluated using a simulation. This limited our ability to assess the full potential of the tool. A next step would be a field-test with human users, measuring the actual composition time, the translation results, and the post-editing effort required when using the tool in comparison to using “regular” MT technology.

In future research we plan to investigate further translatability and the semantic relatedness estimations and automatically tune the metric weights. We further wish to improve the user interface to enhance its utility. Specifically we wish to merge phrase suggestions that are substrings of other suggestions in order to present a more compact list thus making the selection faster and more efficient.

References

- Carbonell, J. G., Gallup, S. L., Harris, T. J., Higdon, J. W., Hill, D. A., Hudson, D. C., Nasjleti, D., Rennich, M. L., Andersen, P. M., Bauer, M. M., Busdiecker III, R. F., Hayes, P. J., Huettner, A. K., McLaren, B. M., Nirenburg, I., Riebling, E. H., Schmandt, L. M., Sweet, J. F., Baker, K. L., Brownlow, N. D., Franz, A. M., Holm, S. E., Leavitt, J. R. R., Lonsdale, D. W., Mitamura, T., and Nyberg, r. E. H. (2000). Integrated authoring and translation system.
- Choumane, A., Blanchon, H., and Roisin, C. (2005). Integrating translation services within a structured editor. In *Proceedings of the 2005 ACM symposium on Document engineering*, DocEng '05, pages 165–167, New York, NY, USA. ACM.
- DeNero, J., Gillick, D., Zhang, J., and Klein, D. (2006). Why generative phrase models underperform surface heuristics. In *Proceedings of the Workshop on Statistical Machine Translation*, StatMT '06, pages 31–38, Stroudsburg, PA, USA.
- Dice, L. R. (1945). Measures of the amount of ecologic association between species. *Ecology*, 26(3):297–302.
- Dymetman, M., Lux, V., and Ranta, A. (2000). Xml and multilingual document authoring: Convergent trends. In *COLING*, pages 243–249. Morgan Kaufmann.
- Hu, C., Resnik, P., Kronrod, Y., Eidelman, V., Buzek, O., and Bederson, B. B. (2011). The value of monolingual crowdsourcing in a real-world translation scenario: simulation using haitian creole emergency sms messages. In *Proceedings of the Sixth Workshop on Statistical Machine Translation*, WMT '11, pages 399–404, Stroudsburg, PA, USA.
- Koehn, P. (2010). Enabling monolingual translators: Post-editing vs. options. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 537–545, Los Angeles, California. Association for Computational Linguistics.
- Koehn, P. and Haddow, B. (2009). Interactive assistance to human translators using statistical machine translation methods. In *Proceedings of MT Summit XII*, Ottawa, Canada.
- Moore, R. and Quirk, C. (2007). An iteratively-trained segmentation-free phrase translation model for statistical machine translation. In *Proceedings of the Second Workshop on Statistical Machine Translation*, pages 112–119, Prague, Czech Republic. Association for Computational Linguistics.
- Papineni, K., Roukos, S., Ward, T., and Zhu, W.-J. (2002). BLEU: a Method for Automatic Evaluation of Machine Translation. In *Proceedings of ACL*, Philadelphia, Pennsylvania, USA.
- Stolcke, A. (2002). Srlm-an extensible language modeling toolkit. In *Proceedings International Conference on Spoken Language Processing*, pages 257–286.

