Authoring Support for Controlled Language and Machine Translation

A Report from Practice

Melanie Siegel
Hochschule Darmstadt
melanie.siegel@h-da.de

Automatic authoring support for controlled language on the one hand and machine translation on the other have previously been seen as two distinct tools in the text processing process. We describe methods for the close integration of both, resulting in better written documents as well as machine translation output of higher quality. The methods were implemented in an existing tool for automatic authoring support.

1   Introduction

With the internationalization of the market for technology products and technologies, the demand for translation of technical documentation is increasing. Especially in the European Union there is a raising awareness that it is not sufficient to provide English language documentation, but that documentation needs to be translated into the native language of the users. These translations must be quickly available, upgradable, available in multiple languages simultaneously, and of high quality. At the same time there are significant technological advances in machine translation: there are rule-based and statistical systems, but also hybrid translation methods. More and more companies are supporting their translation efforts with machine translation. Nevertheless, there are several problems:

• Users are not familiar with the possibilities and limitations of machine translation. This is why their expectations are not met and they are left disappointed.

• To evaluate and test the systems inappropriate texts, such as prose, are used.

• Technical documentation that is translated using machine translation often lacks sufficient quality comparable to texts that are sent to human translators. However, human translators can compensate for this lack of quality in the source document while machine translation systems cannot.

• Statistical machine translation systems must be trained on parallel data. Often translation memory data files[1] are used for training. However, since this data is often accumulated over a number of years and has been originally translated by various different translators, it contains erroneous or inconsistent translations. The machine translation systems trained on this data reflect this heterogeneity of their training data which consequently leads to translations of bad quality.

Authoring support with language technology methods aims at supporting authors in the writing process. Tools based on methods from computational linguistics such as Acrolinx (www.acrolinx.com), are often used by authors of technical documentation. These tools provide support in checking of spelling,

---

[1] Translation memory data contains human-translated parallel sentences that professional translators store in a database in order to reuse the translations. For a more detailed description, see section 3.3.

grammar, and style, as well as in terminology extraction, terminology checking, and sentence clustering[2]. Users of authoring support software, on the other hand, make use of translation memory tools, such as TRADOS (www.trados.com). These tools provide access to previously translated sentences, based on fuzzy matching algorithms. Users of authoring support tools are currently starting to try out machine translation software. Often enough, they are not aware of the possibilities and restrictions of different machine translation methods and tools, which causes them to give up on these tools altogether. For the aforementioned reasons these tools remain distinct, although users already combine them in their daily work.

With the exception of some experiments such as Roturier (2006), authoring support and machine translation have generally been considered two distinct areas. We want to show that both of these areas can benefit from a combination of methods.

Our first goal is to specify some of the possibilities and limitations of machine translation, to deduce options for the authoring of source language documents and to support these through automated processes.

Our second goal is to reduce the manual post-editing effort by enhancing the quality of the machine translation output, since experiments (such as in Allen 2001 and Plitt and Masselot 2010) have shown that manual post-editing on high-quality machine translation output is much faster than the process being completed entirely by humans.

We combine methods from machine translation and authoring support and thus enhance the writing as well as the translation process. Methods and processes need to interfere, as authors and translators need to understand the other's way of working.

To this end we will give an overview of related work that our methods and ideas build upon. We will introduce relevant methods of authoring support, machine translation, and human translation. We will describe how methods from authoring support help users of machine translation to get better performance while explaining at the same time how methods from machine translation can be used in authoring support.

Experiments and evaluations on which the paper is based have mostly been conducted on German language examples. Rules and methods are implemented using the Acrolinx authoring support software and tested on google translate, Langenscheidt T1 and Systran machine translation systems.

2   Related Work

Banerjee et al. (2012) investigate the effect of normalizing source language text, through, for example, spell checking, in comparison to data supplementation. They conclude that "[f]or more noisy datasets […] normalization does improve translation quality more effectively than data supplementation."

Thurmair (2009) lists several pre-editing methods for statistical machine translation: compounding analysis, syntactic pre-processing (only well-formed phrases are allowed in the phrase table) and reordering of source language words in a sentence. Our approach to pre-editing is similar to Thurmair's, but focuses on pre-editing rather than on pre-processing, and it tries to find a set of rules that are efficient for pre-editing.

In Siegel and Bredenkamp (2011) we have shown the impact of controlled language writing to the understandability and translatability of text, as was also done by Reuter (2003). This research focused on human understanding and

---

[2] For a detailed description, see section 3.1.

translation. Inspired by these ideas, this paper takes a closer look at machine understanding and translation.

Genzel (2010) and others reorder the words in the source language sentence, so that it fits better the word order of the target language. They can show an improvement on the BLEU score of the translation. In contrast, we focus on pre-editing mechanisms that tend to have a positive effect on readability, correctness and consistency of the source language text as well as on machine translation.

Hutchins (2005) gives examples for controlled English rules that might affect machine translation quality. He does not refer to automatic checking of these rules.

Thicke (2011) describes the effect of using rules described in the Global English Style Guide (Kohl 2008) for pre-editing and terminology adaptations in the machine translation system: "untrained MT is two times faster to post-edit than to translate from scratch; trained MT is three times faster; and trained MT with source control is four times faster." We follow a similar approach to pre-editing, but also include rules that are specific for machine translation, while the Global English Style Guide was assumed for human translation and non-native English.

Various studies, including O'Brien and Roturier (2007) and Aikawa et al. (2007) have shown that pre-editing of the source language text could lead to machine translation quality improvements (either in terms of comprehensibility or post-editing efficiencies). These studies already take automatic rules into account.

Plitt and Masselot (2010) evaluate the productivity of using statistical machine translation and post-editing in comparison to human translation. They found that "the post-editing of statistical machine translation […] allows translators to substantially increase their productivity." Even more surprising was the result of the quality check: "To our surprise, translation jobs contained a higher number of mistakes than post-editing jobs." The use of machine translation therefore did not only lead to productivity increase, but also to better translation quality. We go one step further and try to support productivity and translation quality with automatic authoring support in the post-editing process.

Simard et al. (2007) describe an automatic post-editing method, in which a rule-based machine translation system is enhanced with statistical post-editing. This method copes for translation errors concerning terminology. This approach is similar to our idea to use statistical machine translation methods for multilingual terminology extraction and to use the results for terminology checking on the machine translation output in the post-editing phase. However, we propose to leave the post-editing process in the hands of human experts while supporting them with necessary information from the authoring support tool.

3   Basic Methods in Authoring Support, Machine Translation, and Human Translation

In this section, we describe the methods in authoring support, machine translation and human translation that are relevant to our approach. Experiments, implementations and evaluations are based on these methods.

3.1   Methods in Authoring Support

Examples for linguistically-based authoring support are Acrolinx (www.acrolinx.com) and LanguageTool (www.languagetool.org). These systems first analyse the input using language technology methods, such as tokenization, morphology analysis and part-of-speech tagging. This linguistic annotation of language data is the basis for further processing steps. We implemented the ideas

described in this paper as part of the Acrolinx software and were thus able to build on existing language technology in this tool.

Spell checking and grammar checking can be based on morphology information and rules that detect spelling and grammar errors. Thus, rules that require context information can be implemented. For example:
"Meine Muttersprache ist Englisch." (My native language is English) vs. „Das Auto ist englisch." (This car is English.)

The words "Englisch" and "englisch" are written with a capital or a small case letter, depending on the context they are used in, and on their syntactic category (POS): it is a noun in the first case and an adjective in the second.

Using the same mechanisms, **style rules** are defined. Style rules mark language constructions that are not wrong as such, but difficult to understand or not exactly adequate for the respective text type. For example, passive constructions can be hard to understand and should not be used in technical documentation. Kohl (2008) describes style rules for Global English and their implementation in authoring support software.

Closely connected to spell checking is **terminology checking**. Again, while spell checking marks errors, terminology checking marks words that are not adequate for the text in the domain. The users (who generally work in a group of authors) define a list of terms in a term bank, which are important for the domain and the texts. They also define a list of words that should not be used in that domain. Connecting this user-given information to the general language analysis means that inflectional variants of these terms can be found and marked, such as plural forms. A term variant detection algorithm makes sure that further variants (such as writing a compound with and without a hyphen) are marked as deprecated. In order to set up such a term bank, the user is supported by terminology extraction methods. The aim of terminology extraction is to automatically detect domain terminology in a given corpus. Terminology extraction is carried out using rules that build upon POS information and lemmatization that makes use of morphology information.

The last relevant method is **sentence clustering**. The authoring support is able to analyze a large amount of text data and find formulation variants, such as "Stellen Sie die Maschine jetzt an." and "Stellen Sie jetzt die Maschine an." (turn on the machine). Here, the word order is distinct but there is no difference in meaning. The user gets the suggestion to always pick one of these variants in order to make the documents more consistent and therefore better translatable.

3.2   Methods in Machine Translation

First we look at two basically distinct approaches to machine translation: the statistical approach and the knowledge-based (or: rule-based) approach. There are also quite a few attempts to combine the ideas of both, as in Eisele (2009) and in Thurmair (2009), but a clear distinction at this point makes it easier to evaluate the influence authoring support has on each of the two methods. Some of the processes we propose can contribute to a hybrid machine translation system.

The **knowledge-based approach** to machine translation makes use of linguistic information and dictionaries and defines translation rules based on this information. Examples for such systems are Lucy (www.lucysoftware.com), Langenscheidt T1 (www.langenscheidt.de), and SYSTRAN (www.systransoft.com). We expect an effect on the machine translation output when source language texts are more correct and therefore easier to analyze with language rules.

The **statistical approach** to machine translation analyzes large amounts of parallel data, sets up phrase tables in the source and in the target language and combines the phrases in the translation task. This kind of parallel data is, for example, available in

translation memories. Examples for this approach are Google Translate (translate.google.de) and Moses (Koehn et al. 2007). We expect an effect on the machine translation output, when source texts are standardized, giving less variation in formulations of text and training data.

3.3   Methods in Human Translation

Most technical translations are carried out with the help of translation memories. **Translation memories** are databases of sentences that have been professionally translated in previous work. While translating, the translator uses the translation memory database that tries to find sentences similar to the one he or she is about to translate. If a similar sentence is found in the database, it can be re-used in the new translation. Somers (2004) shows that the technology of translation memories is similar to example-based translation – a method that is essential for statistical machine translation which is in widespread use today. This is why it can be said that many translators already use machine translation technology today when using translation memories without realizing that they do.

The databases that gradually develop from the work of translators are a valuable data source for statistical machine translation. They provide parallel sentences, translated by professional translators.

4   Authoring Support for Better Machine Translation Results

The language analysis technologies of intelligent authoring support tools provide a basis for improving machine translation: automatic authoring support takes over the tasks of tokenization, POS-tagging, morphology analysis, decomposition and shallow grammar analysis.

Authoring support provides rules concerning monolingual texts. In the process of translation these rules can be applied to the source language text (pre-editing) and to the target language text (post-editing). Further, authoring support rules and methods can help in the evaluation of machine translation results.

4.1   Pre-Editing: Optimization of Source Language Documents

Often enough, users test machine translation engines using texts that are not suited for that task. Sometimes they run tests with prose texts, such as excerpts from historical literature. Prose typically contains a lot of metaphorical language that is difficult to translate even to human translators. Sometimes they process texts of low quality, full of errors.

Using authoring support in the pre-editing process means to correct spelling and grammar errors in the source document. This example of a translation (of an instructive text) using Langenscheidt T1 demonstrates the effect on data cleaning:

„Achten Sie darauf das der LNB fest am Arm des Spiegels montiert ist.“ – „You pay attention to that that the LNB firmly at the arm of the mirror is mounted.“
     but:
„Achten Sie darauf, dass der LNB fest am Arm des Spiegels montiert ist.“ – „Pay attention to the LNB being firmly mounted at the arm of the mirror.“

Consistent and precise terminology helps the machine translation process. If a terminology database is available, checking for terminological consistency supports the machine translation process. For example, in our experiments, the imprecise term "Kerzenschlüssel" was translated to "candle key" while the precise term "Zündkerzenschlüssel" was correctly translated to "sparking-plug wrench". Authoring support provides not only methods for terminology checking, but also methods for terminology extraction in order to set up a terminology database, as

described in section 3.1. Terminology extraction rules in software like Acrolinx are based on linguistic information and run on data in the relevant domain. Thus, the extracted terms are more useful than, for example, a general domain-independent ontology.

Authoring support also contains style rules. These are defined along the lines of consistency, clarity, understandability and translatability of text. Many of these rules are also useful for machine translation pre-editing. Here is an example of a complex sentence structure, where the general authoring rule for reformulation is also helpful for machine translation:

"Eine zu lose Zündkerze hat Wärmestau und schlechte Abdichtung zur Folge, ein zu kräftiges Anziehen hingegen kann das Kerzengewinde und sogar den Zylinderkopf beschädigen." – „A too loose spark plug has been able to damage accumulation of heat and bad seal for the consequence a too strong pickup on the other hand the candle thread and even the cylinder head."

but:

„Eine zu lockere Zündkerze führt zu Wärmestau und schlechter Abdichtung. Eine zu fest angezogene Zündkerze kann das Gewinde und den Zylinderkopf beschädigen." – „A too loose spark plug leads to accumulation of heat and bad seal. A too firmly absorbed spark plug can damage the thread and the cylinder head."

Another area of authoring support that can be of help in machine translation is sentence normalization. Given a large amount of text data in the domain, sentence clustering can find similar sentences and help standardize them. Here is an example of an English sentence cluster found in technical documentation:

- For more information, consult our web page.
- For more information, consult the web page.
- Go to our web page to find more info.

Authoring support can make sure that in contexts like this only one specific variant is used. If the variant selection is trained on machine translation training data, it can be made sure that in case of variants in the source language text these are changed to 100% matches in the machine translation training data.

These methods of pre-editing can, on the one hand, be applied by authors, as is usually done in the technical documentation authoring process. Errors are marked so that the author can come up with reformulations. Further, the author gets a better understanding of the abilities and limits of machine translation as such.

On the other hand, it is possible to automatically apply many of these reformulations. In contrast to authoring support of technical documents, the main focus here is on better machine translation results. Automatic application of rules is much faster. It can be the case, however, that the writing style used in the source language text gets deteriorated when automatically applying rules without any human intervention whatsoever.

4.2   Post-Editing: Optimization of Machine Translation Output

The authoring support tool works on and processes monolingual text. Therefore, using the same mechanisms as in pre-editing, we can correct spelling, grammar and style errors on the target language text. In order to collect errors to correct in automatic post-editing, we conducted experiments with professional translators performing post-editing on machine translation output. On top of spelling and grammar corrections, we identified rules for automatic post-editing of German machine translation output on this basis. Here are examples of these rules:
- correct terminology
- correct standard expressions

- correct word order
- convert future tense to present tense
- convert indicative to causative
- convert "man" to passive
- convert series of "von"+ noun

Terminology correction is a major part of the corrected errors and thus requires a specific focus. Allen (2001) shows tools that support humans in manually adding unknown words to a dictionary and in domain-adaptation by manually selecting the best translations of words in a certain domain. We propose to do multilingual term extraction (a method that is part of authoring support), and make use of term bank organization (with domains and linked terms) for the organization of the domain-dependent selection of translations. The terminology can be adapted to the domain by extraction methods working on the training data; a term bank is set up with information about the domain and used for terminology checking on the source and target language texts.

4.3   Evaluation of Machine Translation

The authoring tool can also be used to evaluate machine translation results. Already in the project Verbmobil (Wahlster 2000) the idea came up to run different machine translation machines in parallel and to choose the best result. What result can be seen as best is dependent on the translation task. In the Verbmobil environment, picking the best result was often enough guided by the time restriction because this was a spoken dialogue translation situation, where processing time is crucial. In translating technical documentation, a few seconds more or less are not that important. More important, however, is the actual quality of the machine translation result. The quality can be determined using authoring support along the areas of spelling, grammar, style and terminology checking of the target language text. Checking results are combined in an evaluation value that can directly be used to choose the best translation.

5   Machine Translation Methods for Better Authoring Support

The benefit of combining authoring support and machine translation is bidirectional because authoring support can also benefit from methods developed for machine translation. The first connection is very basic: some of the rules developed in authoring support for machine translation pre-editing can also be used in authoring support of technical documentation writing. Technical documentation often enough has to be translated to other languages. If the source language document is organized such that it is easier to translate for machine translation systems, it is also easier to translate (and to understand) for humans. Consider the rules concerning reduction of ambiguity here, for example. One of these rules requires avoiding anaphoric pronouns. MT translates single sentences and is mostly unable to take care of anaphora resolution. Human translators use translation memories that also consist of single sentences. Therefore, this rule is useful for both.

Another area is the set-up of a term bank for terminology checking. Using paradigms from statistical machine translation, we extract phrase tables from multilingual data. These phrase tables build the base of linking, on the one hand, terms between languages and, on the other, terms within a language. The result is that we have within each language synonyms of the extracted terms because these terms were translated in the same way. For example, in German – English parallel data we found this cluster of words for English because all of these were translated to

the German word "Grundstellung": starting position, basic position, home position, basic setting, initial position, starting pos., normal position. This word cluster actually builds a synonym cluster. These synonyms can be imported into the term bank and then linked. This information can be used in terminology checking, but also in sentence clustering as a similarity measure.

Finally, the multilingual term extraction results (using statistical machine translation methods) can again be used for rule-based machine translation to enhance the domain-specific dictionaries and for statistical machine translation to make the training data more consistent.

6   Experiments and Evaluations

6.1   First Experiment: Evaluate what Type of Authoring Support is Useful for MT

The goal of the first set of experiments we conducted was to find out how to set up the authoring support system in order to get the best results on machine translation output. We started with only three documents and one rule-based machine translation system, Langenscheidt T1. We applied several existing rules of authoring support and inspected the machine translation behavior. Here spelling and grammar correction on the source text turned out to be an important factor. Further, we found out that rules concerning lexical items, such as the avoidance of ambiguous words, have an important impact on the quality of machine translation.

	Google Translate		SYSTRAN	
	English	Italian	English	Italian
Avoiding *man*	3	3	nn	nn
Separation of verb components	2	2	2	1
Fillers	1	1	3	3
Prepositions	1	1	nn	nn
Word order	3	3	nn	nn
can	3	3	nn	nn
Simple list	nn	nn	nn	nn
Enumerations	3	2	2	2
One relative clause	3	3	3	3
Two relative clauses	3	3	2	2
Passive	3	3	3	2
Conditional clauses	2	2	2	2
Prepositional phrases	2	1	2	1
Questions	2	1	1	1
Swapping subject-object arguments	nn	nn	3	3

3: strong improvement, 2: slight improvement, 1: no improvement, nn: improvement not necessary

Table 1: Application of rules and the effect on Google Translate and SYSTRAN for German-English and German-Italian translation. Adapted from Klausner (2011: 43).

6.2 Second Experiment: Different MT Systems and Target Languages

The next step was to take into account a statistical machine translation system (Google Translate) and a rule-based system (SYSTRAN) in order to find out if pre-editing rules have more influence on one or the other. Furthermore, we compared the translations German – English and German – Italian in order to investigate whether source language correction methods should be dependent on the target language. For the results of this experiment, see Table 1. We identified style rules that have a similar effect on both machine translation systems. These include, for example 'do not place two parts of a verb too far away from each other in the sentence' ("separation of verb components"), or rules that have a bigger impact on the statistical system, such as 'avoid the impersonal pronoun *man'* as well as rules that have an influence on the rule-based-system, such as 'use standard word order'. Simple lists can be handled by both machines and both target languages and need not to be converted, while enumerations seem to be a problem in MT.

In this experiment[3], we could not find any differences in the influence of authoring support rules on machine translation results that can be explained by the structure of English or Italian as a target language. However, the results of the experiment could be used to produce a linguistic resource for authoring support checking, i.e., the 52 style rules used in the following experiments.

	Evaluation (original text)	Evaluation I (pre-edited)	Evaluation II (pre-edited)	Evaluation III (pre-edited)	Average evaluations (pre-edited)
content words wrong	42	81	77	37	65
incorrect punctuation	11	4	8	4	5.33
incorrect word forms	17	27	7	3	12.33
incorrect word order	32	50	30	42	40.67
missing content words	16	1	17	3	7
other error	27 (15.79%)	15	27	19	20.33 (11.15%)
wrong functional words	26	34	40	21	31.67
total	**171**	**212**	**206**	**129**	**182.33**

Table 2: Classification of errors in MT output of original text compared to text, where pre-editing rules are applied

6.3 Third Experiment: Evaluate the Effect of Implemented Rules on MT Results

The aim of the next round of experiments was to evaluate the effect of these rules on machine translation results. To this end, a German text from the Open Office documentation (http://opus.lingfil.uu.se/OpenOffice3.php) was corrected along the lines of the authoring support checking results. Original and corrected text (about 100 sentences each) was translated using Lucy (rule-based), Moses and Google Translate (both statistically-based) and their results classified by translators according to the identified translation errors.

The error classification (see Table 2) showed that the translation of the pre-processed text contained more errors that could be classified by professional technical translators. The category "other error" contained 15.79% for the machine translation

---

[3] The experiment was supported by Katharina Klausner. For further information see Klausner (2011).

without pre-editing and only 11.15% for the machine translation with pre-editing.[4] It could also be shown that the results contained less grammar errors (punctuation, word forms), if pre-editing was performed with the authoring support rules.

Table 3 shows that there were no machine translation results categorized as correct translations for the non-pre-edited texts while with pre-edited machine translation, 20% of the translations were correct.

	Evaluation (original)	Evaluation I (pre-edited)	Evaluation II (pre-edited)	Evaluation III (pre-edited)	Average evaluations (pre-edited)
1.1 Lucy	0	1%	5%	10%	5.33%
1.2 Moses	0	1%	3%	9%	4.33%
Google Translate	0	4%	7%	7%	6%
total	**0**	**7%**	**20%**	**33%**	**20%**

Table 3: Comparison of sentences classified as correct translations, original and pre-edited sentences

6.4   Fourth Experiment: Human Post-Editing of MT Results

In another task (see Table 4), translators were asked to post-edit machine translation results. We calculated the total word-level Levenshtein distances between machine translation output and post-edited machine translation output. Pre-editing using authoring support before machine translation processing lead to the reduction of distance (15.82%) which indicates that post-editing is much less effort when pre-editing was involved as well.

(Original)	2225
I (pre-edited)	2113
II (pre-edited)	1924
III (pre-edited)	1582
average pre-edited	1873

Table 4: Total word-level Levenshtein distances between machine translation output and post-edited machine translation output

6.5   Fifth Experiment: Automatic Evaluation of Rule Impact

In Roturier et al. (2012), we introduced a framework to analyze the impact of source re-formulations on machine translation quality using automatic metrics. This approach will enable us to automatically evaluate the effect of pre-editing rules on machine translation quality and therefore to get evaluations on more data. First results show that grammar reformulations (leading to correct grammar) seem to have a large influence on translation quality.

---

[4] Other errors could be very different, such as untranslated words as in "By clicking of this symbol you add a hyperlink of the current URL to your document <u>ein</u>." or translation close to the (German) source text, such as "Did you split the page into columns or the cursor is in a multi-column text frame, …".
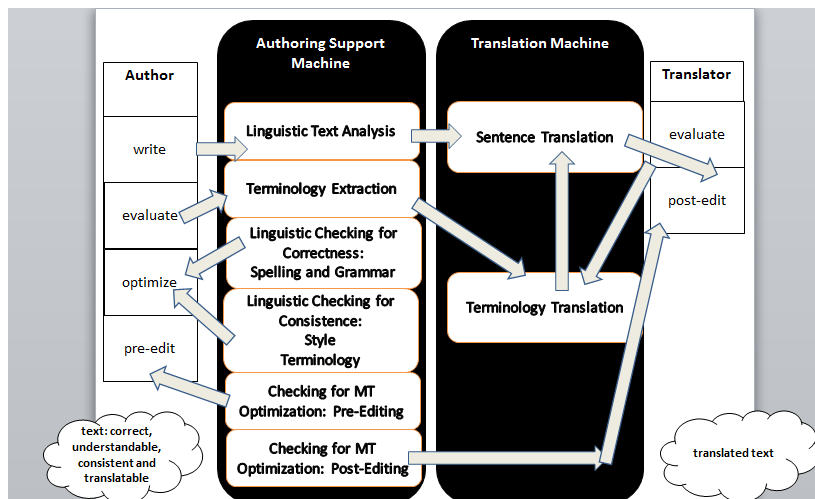
Figure 1: Workflow and Interdependencies in a Combined System

7   Summary

We have shown that methods from automatic authoring support and machine translation can be applied to both tasks. Thus, authoring support and machine translation should be integrated modules in the text production and translation process. Automatic authoring support is monolingual: checking rules can only be applied to either the source language text or the target language text, without taking the translation into account. Therefore, multilingual terminology extraction on SMT training data is a first step towards multilingual checking.

In order to further expand the integration, it is necessary to involve more languages in the evaluation tasks. We plan to include Chinese in our inspections. As for the process itself, we will automate the pre- and post-editing process as much as possible and evaluate whether the machine translation results can be optimized without or with minimal human intervention.

8   Acknowledgment

9   References

Aikawa, T., Schwartz, L., King, R., Corston-Oliver, M., & Lozano, C. 2007. Impact of controlled language on translation quality and post-editing in a statistical machine translation environment. *Proceedings of the MT Summit XI.* Copenhagen, Denmark, 10-14.

Allen, J. 2001. Postediting: an integrated part of a translation software program. In: *Language International*, April 2001, 26-29.

Banerjee, P., Naskar, S.K., Roturier, J., Way, A. and van Genabith, J. 2012. Domain Adaptation in SMT of User-Generated Forum Content Guided by OOV Word Reduction: Normalization and/or Supplementary Data? *Proceedings of EAMT 2012.*

Eisele, A. 2009. Hybrid Architectures for Better Machine Translation. *GSCL Workshop „Kosten und Nutzen von MT"*, Potsdam, September 2009.

Genzel, D. 2010. Automatically learning source-side reordering rules for large scale machine translation. *Proceedings of the 23rd International Conference on Computational Linguistics (COLING '10).* Association for Computational Linguistics, Stroudsburg, PA, USA, 376-384.

Hutchins, J. 2005. Current commercial machine translation systems and computer-based translation tools: system types and their uses. *International Journal of Translation*, Jan-Dec 2005, 17 (1-2), 5-38.

Klausner, K. 2011. *Einsatzmöglichkeiten kontrollierter Sprache zur Verbesserung Maschineller Übersetzung*. Bachelorarbeit, Fachhochschule Potsdam, Januar 2011.

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. and Herbst, E. 2007. Moses: open source toolkit for statistical machine translation. In *ACL 2007, Proceedings of the Interactive Poster and Demonstration Sessions*, Prague, Czech Republic, 177–180.

Kohl, J. R. 2008. *The global English style guide: writing clear, translatable documentation for a global market*. SAS Inst.

O'Brien, S., & Roturier, J. 2007. How portable are controlled language rules? A comparison of two empirical MT studies. *Proceedings of MT Summit XI*, 345-352.

Plitt, M., & Masselot, F. 2010. A Productivity Test of Statistical Machine Translation Post-Editing in a Typical Localisation Context. *The Prague Bulletin of Mathematical Linguistics*, (93), 7-16.

Reuther, U. 2003. Two in One – Can it work? Readability and Translatability by means of Controlled Language. *Proceedings of EAMT-CLAW*.

Roturier, J. 2006. *An Investigation into the Impact of Controlled English Rules on the Comprehensibility, Usefulness and Acceptability of Machine-Translated Technical Documentation for French and German Users*. Unpublished PhD dissertation. Dublin City University.

Roturier, J., Mitchell, L., Grabowski, R. and Siegel, M. 2012. Using Automatic Machine Translation Metrics to Analyze the Impact of Source Reformulations. *Proceedings of AMTA 2012.*

Siegel, M. and Bredenkamp, A. 2011. Localization and writing for the international market. *TC World*, March 2011.

Simard, M., Ueffing, N., Isabelle, P., Kuhn, R. 2007. Rule-based translation with statistical phrase-based post-editing. *ACL 2007 Second Workshop on Statistical Machine Translation*. Prague.

Somers, H., & Diaz, G. F. 2004. Translation memory vs. example-based MT: What is the difference. *International Journal of Translation*, 16(2), 5-33.

Thicke, L. 2011. Improving MT results: a study. *Multilingual*, 22(1), 37.

Thurmair, G. 2009. Comparing different architectures of hybrid machine translation systems. *MT Summit XII: proceedings of the twelfth Machine Translation Summit*, August 26-30, 2009, Ottawa, Ontario, Canada, 340-347.

Wahlster, W. 2000. *Verbmobil: Foundations of Speech-to-Speech Translation*. Springer Verlag, Berlin.