

# Integrating Machine Translation (MT) in the Higher Education of Translators and Technical Writers

**Marion Wittkowsky**

Flensburg University of Applied Sciences

## **ABSTRACT**

This paper describes how MT is integrated into a course project for translation and technical writing students. The course project is based on the idea of combining controlled language and a pre-editing step in order to achieve an effective way to prepare contemporary technical documentation for rule-based machine translation (RBMT). I will explain what I mean by “contemporary” within the context of technical documentation and why this attribute plays an important role within the decision-making process to integrate CL, pre-editing and MT in the course project, which also includes practical exercises for the students. In addition, the reason why RBMT is the MT method chosen within the context of multilingual text production is explained.

## **1. Introduction**

Over the last decade the knowledge and skills required for a career as a professional translator or a technical writer working within the environment of multilingual technical communication have changed drastically. Translators are not only working as translators; they have to manage projects, define the language and terminology to be used, and they have to have knowledge of language, translation and project management software and systems. Today, technical writers not only have to be at least bilingual experts in writing well-structured technical texts, they also have to create style guides and need to know how to optimize information distribution and knowledge transfer. The skills and knowledge that today’s students of technical writing and translation acquire, can be seen as wide-ranging knowledge within the field of international technical communication, i.e. they become international technical communication experts. In this context, I am referring to those students who have been educated in this occupational area over the last 10 years. “Contemporary” technical documentation, no matter whether it is of informative, instructive or descriptive content, is produced to a great extent by technical authors who did not study international technical communication with all its various

aspects. A very high number<sup>1</sup> of persons writing or generating single-language technical contents today are not trained technical writers, but rather engineers or experts in a certain technical area, so-called “career changers”. Why should this be an important factor when it comes to translation processes? Writing technical texts without considering the fact that a subsequent translation process only works properly if the source text follows certain rules, in most cases results in texts that are of course comprehensible and readable yet not necessarily easy to translate, neither for humans with or without for CAT nor for MT systems. As this probably affects more than 70% of the technical documentation that is currently being written, there has to be an approach on how to interact within the translation process so that the source texts meet the language requirements of the different translation resources to be used afterwards.

## 2. Content of the course project

Starting with the properties of natural languages (NL), the students are given an introduction into the RBMT-system to understand how the system works, so that they will be able to experiment with the use of controlled language (CL) as a pre-edit step prior to RBMT.

### 2.1. Natural languages

Natural languages have certain properties that may cause misunderstandings among the recipients regarding the intended message sent by the sender. One of the main problems with natural languages is ambiguities. Lexical ambiguities, for example, can easily be solved by defining unique terms to be used in a specific subject area. Other types of ambiguities such as syntactical, semantic or contextual ambiguities need to be revised and changed. In addition to language variants that should be avoided, various language styles may lead to a lack of understanding for the target audience.

### 2.2. The MT system

This section discusses why RBMT is used and why this RBMT system is Lucy LT.

Statistical machine translation (SMT) systems work with huge databases to look up whether the desired contents can be found anywhere in the databases. Nowadays, because they have been fed for a long time, SMT systems do find translations for sentences like: German “Die nächste Schraube an vorderster Stelle einsetzen.” You get a translation that reads: “Insert the screw next to the forefront”.

Why did I choose this example? I didn’t choose it, but simply entered a meaningless, yet grammatically correct sentence in Google translate<sup>2</sup>. The system took all the words, looked them up and the result is the above.

What does “nächste” in German mean? In different context it can have the meaning of 1. *next*; 2. *next to* or 3. *very near*.

---

<sup>1</sup> More than 70% of the employees working in technical documentation departments have no formal educational training in technical writing. ([http://www.tekom.de/fileadmin/Dokumente/de/2013-08\\_Branchenkennzahlen\\_2013\\_DE.pdf](http://www.tekom.de/fileadmin/Dokumente/de/2013-08_Branchenkennzahlen_2013_DE.pdf))

<sup>2</sup> Google translate is a service by Google Inc. providing translations for written text in many language combinations.

However, the system translated "nächste" with "next to" (the forefront). From a grammatical point of view, it is not even possible to translate it this way, because "nächste" in German with the meaning "next to" requires an indirect object, which does not exist in the sentence.

This is only one example that shows why SMT is not useful to obtain proper translation results. There are no linguistic rules programmed in SMT systems – that is the reason why we have chosen to use RBMT for our course project.

The MT system used in this course project is Lucy LT. Lucy LT is a rule-based MT system with a long history. The system's origin is METAL, a system that had been developed by the Linguistic Research Center of the University of Texas, Austin. *"It was based on the interlingua philosophy and was also influenced by the then current theories of Transformational Grammar. A system for German-English translation based on these ideas was implemented during the years 1972-1975."* (Whitelock, Kilby 1995: 171)

For our purpose to work with the language pair German – English the METAL system delivered the best working base; it was used until the end of the millennium at the Flensburg University of Applied Sciences. Lucy LT is the successor of METAL and the decision was made some years ago to continue using this system for course work with the students. A very helpful feature within the program is the so-called Scratchpad, where one can translate e.g. single sentences and display the analysis and transfer tree of the source and target sentence to understand how the translation process is performed by the system with regard to syntax rules, etc.

### **2.3. Controlled language**

Technical documentation today often requires that authors need to be questioned regarding the original intended meaning of their texts. In order to find a way to obtain a certain quality level, efforts have been made to define rules on how to use language, avoid ambiguities and achieve a high quality level both in the source and in the target texts. (cf. Babych, Hartley, & Sharoff, 2009) There are guidelines on how to write rule-based texts in German that have been published by associations like tekomp<sup>3</sup> or one can use software like Acrolinx<sup>4</sup> to check the correct application of controlled language.

Technical writers are asked to apply CL to avoid ambiguities, inconsistencies, redundancies, and other problems that may cause problems in understanding. However, we are currently very far away from the point that the majority of technical writers apply these rules, as mentioned above. The future experts of international technical communication need to know what results can be achieved by applying CL, not only for the original texts but also for succeeding translations.

## **3. Course project scenario**

Let me first introduce the target audience, i.e. the students who take part in this course. These students who are studying in this master programme International Technical

---

<sup>3</sup> European Association for Technical Communication – tekomp Europe e.V.

<sup>4</sup> Acrolinx is a language checking program distributed by Acrolinx GmbH, Germany.

Communication have quite varying backgrounds regarding their general knowledge of multilingual text production and translation processes as well as in their professional experience level – we have translators as well as technical writers, and the so-called career changers as the programme can also be studied part-time by people who are already working, either as employees or freelancers. What the students are supposed to gain from this course is not only the knowledge about how MT systems function but also to gain a deeper knowledge of why certain properties of NL can be a problem in the source and in the target texts resp. for the source and target text audience, and how the use of CL can help solve those problems.

### Methodology

The course guides the students through the following steps:

- a. After an *introduction into MT systems* in general and a short briefing about *NL and CL theories*, the students receive instructive and descriptive excerpts from a contemporary technical text<sup>5</sup> (service manual or similar) in order to execute a first machine translation using Lucy LT. The subject area the technical text is taken from is not of special interest because the dictionaries provide a broad variety of technical subject areas.
- b. Then the first actual work done by the students is *entering entries into the dictionary* either by replacing or adding wrong or missing terminology within the monolingual, bilingual and transfer lexicons. Here the students enter the canonical forms, make simple lexical additions to the new terminology, e.g. by entering the word stem, affixes, etc. or specifying the lexical category as for example AST, NST or VST for adjective, noun or verb stem. Because of the fact that we do not educate computer linguists, the work with the dictionaries should not superimpose the actual subject matters of this course.
- c. After having finished the dictionary work, the students are instructed to perform another translation and to view and analyze the results. In many cases the resulting texts at this point are not comprehensible for persons who do not have the same context as the students. Some sentences are perfect and some are completely incomprehensible. The Scratchpad in Lucy allows the students to translate small segments of the source text and to display the analysis tree and the transfer tree where the system behavior can be followed.
- d. Now, the main *experimenting* part begins. As mentioned above, the quality level of the translated contemporary technical texts reached so far is often poor so that the students now start to apply CL rules combining style and grammar to the source texts in order to achieve higher quality target texts. They shorten sentences, simplify the syntax, find solutions for various kinds of syntactical ambiguities, try out the system behavior when using transitive and intransitive verbs, avoid complex phrases and tenses, eliminate redundancies, change the grammatical gender, etc. (tekomp: 2011). This experimenting step is very important for the following reasons:

---

<sup>5</sup> In the first part of the course, it is a German text, in the second part the source text is English.

- (1) The students reduce their distance to the software in a positive way. Observations over the last years working with students have shown that they normally have great respect for computers and software.
- (2) They learn that they are able to make changes and to interact with a computer.
- (3) The students see that changing things may bring about better results.

#### 4. The course project in practice

In this section I would like to show examples of the practical work performed by the students. The language combination relevant in this course project is German - English. As mentioned before, the students focus on modifying a given source text using CL rules in order to improve the translatability of source texts. The students work with the RBMT system Lucy LT, however this work could also be done using any other RBMT system.

Let us start with step 1, the *introduction* to MT systems, a quick overview of traps of natural language and the start guide for using CL. This introduction is important to ensure that all students have the same knowledge level when starting the actual project.

The students get a copy of the original technical text file, we look at the contents and answer possible questions regarding e.g. the subject matter. After having prepared the MT system for the translation such as specifying the subject area, or making some basic language pair settings, or define display options, the students then let the MT system translate the original source text without modifying it in advance. The sentence used in the following section has been taken from an automotive manual but from a linguistic point of view it could be part of any contemporary technical documentation and serves as an example only.

Example source sentence (DE):

*Der Gangwechsel wird mit den Schaltwippen oder dem Wählhebel vollzogen.*

(Meaning: You have two possibilities to change gears: use the paddles or the gear selector.)

1st translation (EN) by Lucy LT:

*The gear trimmer is carried out with the switching rocker plates or the choosing-lever.*

We find that we have three apparently problematic terminology components within this text:

- Gangwechsel ≠ gear trimmer
- Schaltwippen ≠ switching rocker plates
- Wählhebel ≠ choosing lever

The three terms (combinations of two words each) are new to the system, which proposes the usage of two independent target terms as a combination of single words that have been stored in the system.

The result of the first translation of the correct and understandable source sentence makes no sense.

Step 2 involves *entering the* terminology into the three *dictionaries*, two monolingual lexicons and one transfer lexicon. In these dictionaries the students enter the canonical forms as well as basic grammatical information like the lexical category (in this case NST for noun stem) for every term. For our sentence, in detail this means the following:

*Schaltwippen* is a plural word so that the singular word entered is *Schaltwippe* with *paddle* as target.

*Wählhebel* can be entered into the dictionary as it is with *gear selector* as target.

The third term that has not been translated properly by the system is *Gangwechsel*, however it should not be entered into the system, because this word as compared to the two words above is not a concrete but an abstract compound noun made up of *Gang* and *wechseln*. In English the corresponding words are *gear* and *shift*. This problem has to be solved in a later step.

At the end of this step, the students have added or replaced two terms in all three dictionaries.

In the 3<sup>rd</sup> step the students *translate* the same source text again:

*Der Gangwechsel wird mit den Schaltwippen oder dem Wählhebel vollzogen.*

The system applies the newly entered terminology. The output is as follows:

*The gear trimmer is carried out with the paddles or the gear selector.*

We now have the correct terminology for the two changed words, yet we still have the problem with the word *gear trimmer* in English.

At this point the *experimenting step* starts. The students perform the following substeps as often as they find possibilities to make amendments to the source text without changing its intended meaning:

- 1) They enter the changed source text sentence by sentence again and again into the source text window of Scratchpad.
- 2) They have Lucy LT translate the text.
- 3) When the sentence<sup>6</sup> has been translated, the analysis tree and the transfer tree may be used as support, when the students analyze the system behavior and their source and target texts. The students can write down all important observations, copy useful and less useful examples of the source and the target texts as well as the trees into their working files so that they have something to work with and to refer to in their course reports.

The following section shows how our example sentence could be changed, i.e. pre-edited and what the results of the translations are. I speak about pre-editing here and not about applying CL, because the students are not asked to apply language rules that follow CL only but

---

<sup>6</sup> It is also possible to enter more than one sentence into Scratchpad, yet the number of characters is restricted.

they may also change the source to that extent that the grammar is still correct and the original meaning is retained. For every source and target sentence listed below you can read some comments referring to the meaning, language and quality that have been made during the translation procedure.

Since it is the third change + translation, I name it correspondingly:

3<sup>rd</sup> source

*Vollziehen Sie den Wechsel der Gänge, indem Sie die Schaltwippen oder den Wählhebel verwenden.*

The German term *Gangwechsel* needs to be separated into two words as mentioned above, yet we still have two nouns to be able to keep the original verb *vollzogen* (canonical form: *vollziehen*). In addition, the passive voice has been changed to active in such a way that the user is addressed directly using *Sie* (EN you). Regarding the quality, the sentence does not follow the rules of CL, it is very complicated.

3<sup>rd</sup> target

*Carry out the trimmer of the gears when you use the paddles or the gear selector.*

The separation of *Gangwechsel* into *Wechseln der Gänge*, did not improve the translation at all. Lucy still translates this part of the sentence with "trimmer of the gears". This is a mistake.

4<sup>th</sup> source

*Zum Wechseln der Gänge verwenden Sie die Schaltwippen oder den Wählhebel.*

Here, the problematic word *Gangwechsel* has again been separated, yet we have a nominalized verb *Wechseln* and the plural form of *Gang*->*Gänge*, which results in *Wechseln der Gänge*, but which means the same thing. We avoid the verb *vollziehen* because it is not the correct style in this context and use the verb *verwenden* instead.

4<sup>th</sup> target

*You use the paddles or the gear selector to the change of the gears.*

The resulting sentence is not correct regarding the grammar, and the expression *change* could be misunderstood.

5<sup>th</sup> source

*Sie verwenden die Schaltwippen oder den Wählhebel zum Wechseln der Gänge.*

We have the same content and the same words as in example 4, however, the sentence structure has been changed. The sentence does not follow the rules of CL, i.e. one should be informed about the result before taking an action (tekom: 2011, p. 55).

5<sup>th</sup> target

*You use the paddles or the gear selector to the change of the gears.*

Very interesting here, the target sentence is the same as in example 4 and could still be misunderstood.

6<sup>th</sup> source

*Wechseln Sie die Gänge, indem Sie die Schaltwippen oder den Wählhebel verwenden*

The sentence in German is correct and written by applying CL rules. The noun *Wechseln* has been changed to *Wechseln Sie* (verb+addressing the user).

6<sup>th</sup> target

*Change the gears when you use the paddles or the gear selector.*

The system response is correct by changing the noun *change* to a verb, yet the conjunction *when* is wrong – it should be *by* + gerund.

7<sup>th</sup> source

*Sie können mit den Schaltwippen oder mit dem Wählhebel die Gänge wechseln.*

In this example we again do not have the right order regarding action and objective and we have a modal verb that should be avoided (tekom: 2011, p.62).

7<sup>th</sup> target

*You can change the gears with the paddles or with the gear selector.*

Apart from the modal verb, the fact that *with* should be replaced with *using* and the second appearance could be deleted; the sentence is of an acceptable quality.

8<sup>th</sup> source

*Um die Gänge zu wechseln, die Schaltwippen oder den Wählhebel verwenden.*

We have a correct German sentence following CL rules.

8<sup>th</sup> target

*In order to change the gears, use the paddles or the gear selector.*

The resulting target sentence is perfect.

9<sup>th</sup> source

*Gänge wechseln mit Schaltwippe oder Wählhebel.*



Last but not least a very short sentence that summarizes the information and could be used as a title for purely informative purposes.

9<sup>th</sup> target

*Change gears with paddle or gear selector.*

The translation depicts the same as the German source.

There are at least 10 other possibilities to write the intended content in other words. These examples represent an excerpt only.

## 5. Conclusion

The course project presented is intended to make the students aware of other translation scenarios than human translation or computer-aided translation CAT. It is supposed to introduce the students into machine translation, let them form their own informed opinion about MT by showing them how the system behaves and how an advanced human interaction may increase the quality of the outcome. Since the students are to be regarded as experts in international technical communication, they should not only have an interest in the translation process but also in the process of generating source texts of a certain quality level – depending on various factors like target audience, text type etc.

From my point of view it is a valuable side-effect that the students generate different quality levels when experimenting with the texts, so that future research in this area could probably bring up an approach to define different quality levels and to find better ways to generate multilingual text units from existing documentation on demand.

## References

- Babych Bogdan, Anthony Hartley, Serge Sharoff (2009): "[Evaluation-guided pre-editing of source text: improving MT-tractability of light verb constructions](#)". *EAMT-2009: Proceedings of the 13th Annual Conference of the European Association for Machine Translation*, Lluís Màrquez and Harold Somers ed., 14-15 May 2009, Universitat Politècnica de Catalunya, Barcelona, Spain; 36-43
- tekom (2011): *Leitlinie für regelbasiertes Schreiben - Deutsch für die Technische Kommunikation*, Gesellschaft f. technische Kommunikation e.V., tekom
- Whitlock Peter, Kieran Kilby (1995): *Linguistic and Computational Techniques in Machine Translation System Design*, 2<sup>nd</sup> Edition; (Studies in Computational Linguistics) London: UCL press