

The Quest for Machine Translation Quality at CLS Communication

Hans-Udo Stadler, Ursula Peter-Spöndli

CLS Communication AG
Elisabethenanlage 11, 4051 Basel
Switzerland

hans-udo.stadler@cls-communication.com, ursula.peter@cls-communication.com

Abstract

While Machine Translation (MT) quality is usually interpreted as the linguistic quality of the translations themselves, there are other aspects that should be considered, as well. These aspects may depend on the specific system or provider, as there are various types of applications for MT. CLS Communication as a commercial MT service provider has to meet its clients' needs and set itself apart from competitors. We have established a simple and flexible quality management approach that allows us to measure and improve MT quality on an ongoing basis. From a practitioner's point of view, we show how we collect input for the further development of our offer and implement the respective findings in our environment. Only few resources are needed for this approach, and it can easily be adapted to our future requirements.

1 Introduction

As a globally active language services provider specialized in translating, editing and processing multilingual content for the finance, insurance and life sciences industries, CLS Communication also offers Machine Translation (MT) as an online service to its clients. CLS Machine Translation is mainly offered as a "multilingual communication enhancement" tool, enabling employees e.g. to translate foreign-language mails or meeting minutes they receive from a colleague into their mother tongue. Thus, the MT team at CLS has to make sure that the raw translations of texts of many different kinds and from a variety of domains are understandable without post-editing and that the handling meets the users' needs. Quality is an incentive for our users to prefer CLS MT to Internet tools.

2 From Translation Quality to Machine Translation Quality

When people talk about quality, they seem to have a common understanding of what quality means, even if they have rarely ever thought about its definition. General definitions often use the notion of "fulfilling requirements" or the "degree of excellence", which raise questions like: Who specifies the requirements and how can the degree be measured? When it comes to MT systems in particular, the issue of quality usually focuses on the linguistic quality of the results they deliver. In its early days, MT was intended as a fast and cheap substitute for human translators, and Fully Automatic High Quality Translation (FAHQT) was the ambitious but admittedly unobtainable goal (Bar-Hillel, 1964). Over the years, the attitude towards MT has become more realistic and there has been a shift in use as well as in perception. Today, Fully Automatic Useful Translation (FAUT), as propagated by the Translation Automation User Society (TAUS), seems to express the current demand (Joscelyne & van der Meer, 2006). At first glance, the term FAUT looks like a downgrade of FAHQT where "useful" has to suffice when "high quality" is not achievable. But it actually broadens the scope rather than just reducing the value proposition. At the end of the day, what an MT

provider really offers is not just the translations produced by a machine but a bundle of services, so various aspects of suitability or "usefulness" have to be considered. The quality of Machine Translation is reflected in a combination of many things, or, according to ISO, can be seen as the "totality of features and characteristics of a product or service that bear on its ability to satisfy stated or implied needs" (ISO 8402, 1994). While such standards are typically used to compare products or services with each other, they also help to envision the basic essentials of a particular system, in the first place, because every MT system or service is unique in the combination of its features and characteristics and also has to meet different needs. At CLS, we consequently used the concept of quality with the following aspects:

- build an inventory of quality criteria specific to our offering,
- find ways of measuring our quality
- define actions to improve quality, and ultimately
- position our solution accordingly

There are of course already several evaluation models, which work with clearly defined parameters and are very elaborate (cf. Hovy, King & Popescu-Belis, 2002). But, again, they are usually designed to compare different systems with each other or to help the developers build and test new software. We want to efficiently manage the quality on an ongoing basis without using too many resources. We have developed a simple approach that is effective for our own environment. It is dynamic in the sense that it tolerates changes in conditions without having to rearrange the basic method.

3 Views on Quality

CLS Machine Translation was designed as an Application Service Provider (ASP) model based on a third-party system and offering added value in the shape of vast domain-specific lexicons, data security and a service package. In order to find the best third-party system we ran an evaluation on commercially available MT software. (1) We specified the evaluation criteria, taking into account our targeted business model, our technical infrastructure and our targeted user group (Maier & Clarke, 2001). In this context, one must differentiate

between the client and the end users: (2) The end user will be more demanding if the MT costs are charged to his individual account instead of to a (anonymous) corporate cost centre. (3) Of course, the client will monitor the user satisfaction over time, so we cannot offer them a service that will not be used. (4) The group of end users can be further divided into subgroups, for instance: casual users and power users, or even lay users and experts¹, who each have different expectations. (5) At the same time, we have to deal with various stakeholders on the client's side, e.g. staff from procurement, corporate communications or IT security, who all have their own requirements. In short, as an MT service provider we have to consider a range of views on quality. Over the past few years, the main input sources for quality criteria have turned out to be:

- a. Talks with prospects
- b. Talks with clients and end users
- c. Informal feedback from end users (telephone calls, e-mails to our support box etc.)
- d. Formal/structured feedback from end users via surveys
- e. Monitoring of actual usage (text types and lengths, file formats, translation directions etc.)

4 Breaking down Quality – Quality Criteria

Via the channels mentioned above we have identified the following aspects of overall quality, i.e. areas upon which our target groups obviously judge our MT system:

- a. Translation quality
- b. Usability
- c. Translation directions
- d. Supported file formats
- e. Performance

While translation quality forms an area of its own in that it is very linguistically motivated, the other criteria are of a more technical nature and may thus be seen as application issues. In order to keep our quality management approach simple and open for the future, we do not define rigid categories, sub-characteristics or fine-granular criteria but deal with the aspects rather pragmatically. In the following we demonstrate how quality can be interpreted from this perspective. We give a few examples of how it can be measured and improved.

4.1 Components of quality

4.1.1 Translation quality

The fact that MT systems cannot achieve human quality requires a definition of translation quality restricted to machine translation systems. As the provider and especially as the dictionary builder of CLS Machine Translation, we are concerned daily with the boundaries and possibilities of the CLS Machine Translation system.

It is difficult to find a universally-accepted definition for MT quality. In this context, MELBY speaks of "appropriate" translation instead of "good" translation (1997). According to this definition, machine translation

quality is measured according to its purpose and not according to the criteria of human-generated translation. If the purpose of a translation is a fast, indicative translation or a pre-translation for a human reviser, MT is definitely THE appropriate translation. Accuracy and intelligibility need to be taken into consideration as well, i.e. the same kind and amount of information should be available both in the source and the target language (SL and TL, respectively) text, and the grammar rules of the TL should be obeyed (Schwarzl, 2001). The question of adequate style, however, which is a central issue for human translation, becomes obsolete when considering the above mentioned objectives.

How is the above-mentioned definition of "appropriate" quality implemented in our daily work of improving the machine's output? The following example illustrates how our coders must abandon the requirements of human translation on the one hand and on the other, still strive towards the best possible translation. While for a human translator the best translation of a specific term conveys the meaning of the term in the related context perfectly, for an MT dictionary coder doing a translation analysis, the best translation must not be too context-specific. Rather, it must convey the basic or literal meaning of the term even within the range of a single domain. The coder usually has to rate context-specific as less important than domain-specific, for the coder's target term has to be applicable in a different context (see Example 1). To improve the context link-up in future translations, he may determine translation alternatives within the range of a domain. Since translation alternatives are (optionally) displayed to the user, they may improve the translation quality (see Example 2). In any case, the coder's focus is always directed over the rim of his actual context.

| | | | |
|-----------|-------------|---|----------------|
| Human T.: | coagulation | → | blood clotting |
| MT: | coagulation | → | coagulation |

Example 1: French-English medical translation

| | | | |
|------------------------|--------|---|--------|
| General, law, economy: | action | → | action |
| Financial, economy: | action | → | share |
| Financial, economy: | action | → | stock |

Example 2: French-English translation

When compared to human translation, MT has naturally no problems with spelling in the output text. Exactly the opposite applies to typing errors or grammar mistakes in the input text. An incorrect input is likely to render "garbage" output.²

Usually, the end user would like to receive a translation quality that is comparable to human translation even if he does not absolutely need it. His benchmark will always be this almost human level of translation quality.

¹ e.g. translators who use MT and later post-edit the results, cf. Stadler (2005).

² Yet we realised, that some users seem even to benefit from the machine's adherence to orthography, as they use MT as a kind of a spell checker.

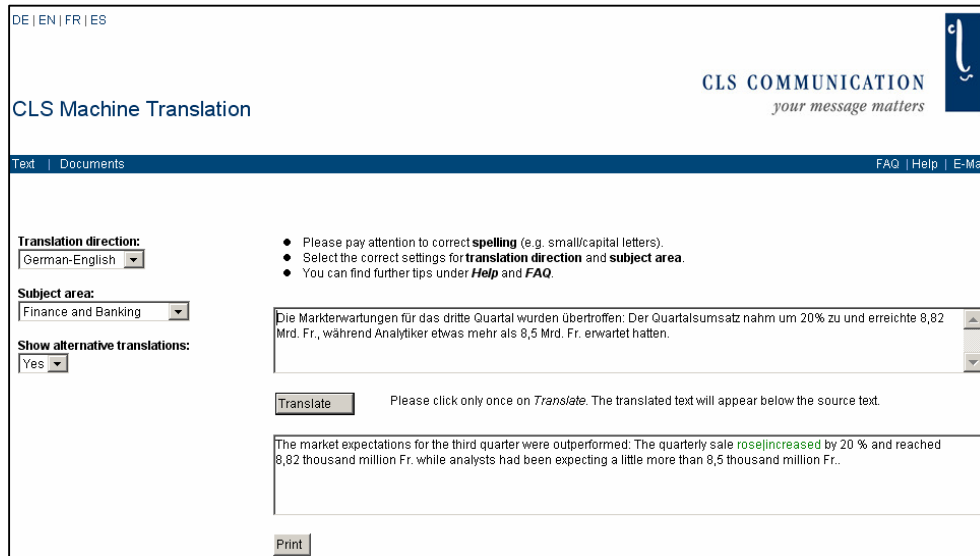


Figure 1: User Interface of CLS Machine Translation

4.1.2 Usability

For our purposes, usability should be understood as the elegance and clarity with which the interaction with CLS MT is designed.³ It thus includes criteria like functionality and layout, but not performance, which we will deal with separately.

In our latest user survey, 99% of the respondents stated that our system was easy to use. However, some of them made suggestions for further improvements, e.g. regarding the selection of translation directions. Currently, the translation direction has to be selected by means of a single drop-down list containing all possible combinations of source and target languages (see Figure 1). So far there are only ten directions available so the list is quite short and requires no scroll bar. Apparently even this list is too long, especially when you always use the same direction and have to select it every time because it is not the first in the list. Some users want a personal profile which stores their individual default translation direction while others would prefer two separate drop-down lists for source and target languages, respectively. The second solution requires two selections by the user, which seems even more complicated at first glance but could be facilitated by an automatic source language detection. We are currently looking into the different possibilities but what these findings actually demonstrate is that people do not want to be bothered with "mundane" administrative tasks. Another item on the wish list was a feedback form enabling the users to give us concrete examples of wrong translations. There is already a button called "E-Mail", which opens a new mail addressed to our support team. Both the low usage of this button and the suggestion of a feedback form imply that the button is either not

prominent enough or it is perceived as being there only for reporting technical problems.

In a broader sense usability starts with data security, even though it might not be recognized as belonging to the usability criteria by the users. Our clients rely on secure transmission and storage of confidential texts, so without the fulfilment of this basic requirement they would not even allow their employees to use an MT service.

4.1.3 Translation directions

Translation directions must be geared to the targeted market. The more translation directions a provider has to offer, the more potential users can be reached. But of course the demand for some directions is lower than for others. In most cases, a particular user needs only one or two language combinations and will thus be satisfied if those are covered. If from a user's perspective one certain direction is missing, he will rate the system's overall quality lower but will still be happy with the existing directions as long as they deliver good results. On a company level, the coverage of important translation directions will be a decisive factor for a client.

4.1.4 Supported File Formats

Apart from plain text segments, CLS Machine Translation currently supports translation of files with the formats .doc, .html, .rtf, and .txt. Usage of these choices is almost completely limited to .doc files. However, some users occasionally try to translate .pdf files, so there seems to be a certain demand for that format. Apparently people do not see why an MT system should not be able to handle such common file formats, so they either do not notice the restrictions (which are mentioned explicitly on our user interface), or they just ignore them and want to see for themselves what works and what doesn't. Other formats like .ppt, .xls or .xml are hardly ever requested by existing end users but would be interesting for specific projects of prospective clients.

³ cf.

http://en.wikipedia.org/wiki/Usability#Usability_considerations

4.1.5 Performance

For the purpose of this quality management approach, performance first of all means response time but also includes availability, both as perceived by the user. Long response times or error messages may result from high load or from problems the engine has with a particular document, and a user cannot always tell the two phenomena apart. Load tests, where usage is only simulated, are helpful in judging the quality but do not represent real-life situations. The same restriction goes for availability, which can be monitored and expressed in percentages, but is only approximate. The most surprising finding of our latest survey was that users had very high expectations towards performance, which some of them felt could not always be met. Browser compatibility is a different issue but can also provoke the notion of bad accessibility on the user's side. When a new browser version or an exotic browser is used it might happen that the MT site can be opened but the text or its translation cannot be transmitted.

4.2 Overall quality

Each of these quality components has to meet the users' requirements to a certain degree in order to guarantee a minimum of overall quality, which has to be monitored and preferably improved over time. As CLS does not work on the software itself but concentrates on lexicon development and the provision of services, in the

following sections we will describe the relevant activities from this perspective.

5 How to Monitor Quality Development

To keep track of the quality development we have to define suitable instruments and a rating scheme.

5.1 Means of monitoring

According to our definition of quality, we currently employ three main instruments to monitor the quality of our MT system.

- a. The most frequently used instrument is **translation analysis**, which is performed for the sake of correcting translation mistakes. It is part of our daily business and produces neither quality scores nor error rates, but serves to detect and correct mistakes in practical terms, as well as to build up dictionary volume.
- b. **Benchmarking** monitors quality on a historical basis. It measures the output quality of the MT system over time by recording changes in quality, be it improvement or deterioration.
- c. While these two methods affect translation quality only, we conduct **surveys** in order to measure other criteria of MT quality, as well, and include the end user's perception of quality.

Table 1 gives an overview of our instruments, which are further described in the following section:

| QE instrument | Measures what | Evaluator | Frequency | Metrics/Criteria |
|----------------------|---|------------|------------------|---|
| Translation analysis | Current translation quality of relevant texts | Coder | Daily | Intelligibility, completeness, appropriateness, style |
| Benchmarking | Change in translation quality of corpus over time | Translator | Quarterly | Changes: improvement or deterioration |
| Surveys | Overall quality of CLS MT | End user | Client-dependent | User satisfaction with regard to translation and application issues |

Table 1: Quality evaluation (QE) instruments at CLS Communication

5.1.1 Translation analysis

The analysis of machine-produced translations stands at the beginning of one of our main coding streams and is an important part of the daily work of our coders. This tool confronts the coder with the output the user may get. The coder has a specific text translated by the machine and analyses it with regard to wrong translations. The sample texts' origins are diverse: They may be texts which were sent to the machine by users before, client texts from the human translations department, or domain relevant texts from the internet. As a general rule, the target language has to be the analyser's mother tongue. Translation analysis is carried out in two steps:

- **Treatment of unknown words:** In a first run, the focus is on words unknown to the machine. The coder sorts out the useful terms and enters them on a list. In a next step he investigates the listed terms' equivalents in the target language. Having completed the research, the terms are added to the dictionary.
- **Correction of translation problems:** Since the analyser can now be sure that syntax problems are not due to missing vocabulary, he then corrects emerging mistakes instantly in the dictionary via the adding or modifying of rules (analysis, transfer, or generation) or lexical features. Since some bugs - particularly syntax

related problems - are due to lingware settings, they are beyond the reach and hence beyond the intervention of the coder. An experienced coder identifies such “hopeless” cases before wasting precious working time on them.

5.1.2 Benchmarking

The benchmark tests cover six translation directions (DE-EN, EN-DE, DE-FR, FR-DE, EN-FR, FR-EN) and are based upon nine texts per direction, each one representing one of the nine subject areas that our automatic translation service provides. The length of the reference texts is roughly one page each.

The benchmark tests are undertaken quarterly. They are performed by one and the same translator who is a member of the coding team. The translation of the last benchmark test three months before serves as reference translation. The classification process is performed at a sentence level. The first criteria decides whether the sentence is translated better or worse than the reference sentence. If the translation has improved, the evaluator assigns the change for the better to the categories “grammar” or “semantics”. If a mistake occurs several times in the same text, it is regarded as one error. Especially in times of large volume imports to our dictionary (e.g. client-specific terminology), the benchmark suite is an important tool to give us feedback about changes in dictionary quality. For our purposes it is more suitable than automated approaches that use reference texts, like BLEU. Changes in the scores triggered by such methods would not really reflect changes in quality but merely changes in translation, which would above all be calculated on the basis of questionable reference texts (human translations).

5.1.3 Surveys

Surveys help us to learn about the end users' needs and understanding of quality. Apart from questions concerning the current usage behaviour, e.g. the use of MT for the translation of specific text types, preferences for certain translation directions or the frequency of use, we also find out about user satisfaction or suggestions for improvement. We place a call for participation on the MT user interface linked to an online survey form. The main findings of our latest survey (Q4 2006) were:

- 80% of the participants use the system at least once a week, one third of them have used it on a daily basis.
- The translation quality met the expectations of almost 80% of the participants.
- Nearly all participants considered the system easy to use and stated that the tool helped them carry out their daily work.
- The majority of the respondents translate between German and English (both directions) as well as from German into French.
- Today, CLS MT is more often used as a pre-translation tool for the subsequent revision of texts, whereas understanding of foreign-language texts was more important in a previous survey. Over the years, it has always been used to look up single words, i.e. as a substitute for dictionaries.

- Most frequently translated text categories included e-mails and internal documents.

5.2 Rating of quality components

Our findings from the described monitoring and the input from the channels mentioned above allow us to roughly rate the current state of CLS Machine Translation quality. For this purpose we employ a scale from 1 to 5, where 1 is low (but, as stated before, high enough to deliver benefits to the users) and 5 is high, leaving room for improvement, of course. Furthermore, the rating can only be tentative, as it would not make much sense to define rigid metrics at one point in time that do not take into account the future state of the art or would define an upper ceiling that could not be exceeded. As the foundation for these ratings we see especially the following considerations, taking into account also the comparison with competitors or free Internet tools:

- Translation quality varies over the different translation directions but our survey showed that it meets the requirements of 80% of our users. We therefore attribute translation quality a rating of 4 out of 5, which is supported by the fact that domain-specific coverage is significantly better than with free Internet tools, for instance.
- Usability was approved by 99% of the survey participants, and data security is guaranteed, as opposed to free Internet tools. We hence rate CLS MT's usability as 5.
- Translation directions are sufficient for most of our existing end users, but some of them would like to have additional directions. On a company level, some prospective clients also inquired about additional language pairs. Some of these languages are offered by competitors and free Internet tools, but so far no single provider covers all directions of interest. We therefore rate this aspect of our system as medium, i.e. 3.
- Supported File Formats: .doc as the probably most widely used format is supported, .pdf still missing. Some competitors can handle .pdf, free Internet tools usually do not support document translation. In this context, we assess CLS MT's quality as 3.
- Performance: Automated monitoring shows good results but, as mentioned above, some users stated that waiting time was sometimes too long, so we attribute a rating of 4.

In conclusion, we assign the following ratings for the current quality of CLS Machine Translation:

| Quality component | Rating |
|------------------------|--------|
| Translation quality | 4 |
| Usability | 5 |
| Translation directions | 3 |
| Supported file formats | 3 |
| Performance | 4 |

Table 2: Rating of Quality Components

6 Quality Management and Product Management

Rating the individual quality components at one point in time delivers only a narrow view as it mainly focuses on the existing deployment of CLS MT. It does not show to what extent the improvement of parts of the system can contribute to the further development of the whole. From a product management point of view we additionally want to consider the future potential, i.e. risks of the quality remaining stagnant, and the opportunities for quality improvements. We therefore assign a weighting to each component, showing where improvement is most important and again using a scale from 1 (low) to 5 (high), based on the following thoughts:

- Translation quality: While we currently rate our translation quality as quite good, we are well aware of the potential threat from competitors working on quality improvement. We also identify additional opportunities resulting from better translation quality like reaching new users, finding new application types and using the system as a basis to increase productivity in human translation. In relation to the other aspects, we therefore weight translation quality with the highest value, i.e. 5.
- Usability: Taking the users' suggestions seriously will further increase their satisfaction. Especially providing them with better feedback options would in return deliver valuable input for further development. Weighting is medium, i.e. 3.
- Translation directions: Additional language pairs bear a certain potential for new clients and additional users. But since our competitors do not cover all relevant languages, either, the risk of losing orders to them is still quite low. We thus weight this quality aspect rather low (2).
- Supported File Formats: Additional formats bear a potential of increased volume and new use cases. But in comparison to the other aspects, they are rather "nice-to-have", thus 2.
- Performance: With ever-faster technology around, users will expect better performance for MT in the future. Neglecting this requirement poses a risk of losing some users. The importance is comparable to usability and is hence weighted 3, as well.

| Quality component | Weighting |
|------------------------|-----------|
| Translation quality | 5 |
| Usability | 3 |
| Translation directions | 2 |
| Supported file formats | 2 |
| Performance | 3 |

Table 3: Weighting of Quality Components

To see what component we should focus on when planning future improvements, we also have to take into account the feasibility, based on the following criteria:

- Costs: e.g. one-time investments (hardware, software etc.) and/or labour costs over time. Costs can be calculated quite reliably in advance.
- Time to market: While the effective hours worked on a particular solution can be expressed as costs, the execution time is another criterion: the sooner the solution can be implemented, the sooner the quality improvement will take effect, resulting in a prompt enhancement of user satisfaction and sales arguments.
- Autonomy: Some solutions can be implemented in-house, while others have to be developed by third-party providers. In-house resources can be allocated according to our own priorities and are therefore relatively easy to manage. Deciding on who to co-operate with also has a strategic dimension.

By combining these three dimensions we rate the overall feasibility of improving each quality component, again using a scale from 1 (low) to 5 (high), see Table 4. For two reasons, it is important to see the quality components in relation to each other, rather than in isolation:

- The dimensions are very heterogeneous and at the same time distributed differently for each of the components.
- Both the possible improvements and the effort can be depicted in absolute values (e.g. costs for a certain number of new translation directions at a certain price) or in relative terms (e.g. progress in translation quality, which is very labour-intensive but hard to measure).

The concepts of current rating, weighting and feasibility all have an impact on the further development. Translation quality, for instance, receives a higher weighting (5) than usability (3) because it opens up much better possibilities (more users and clients, higher translation volume) and is even crucial to our USP. In terms of feasibility, on the other hand, usability gets a higher value (4) than translation quality (3) because some of the suggestions that our users made are very easy to implement (at low cost and within a short period of time), while the improvement of translation quality requires more resources, though still less than new translation directions (feasibility = 1).

By multiplying the values of weighting and feasibility we then see the improvement potential, i.e. how efficiently each issue can be addressed in theory. If an issue already has a high current rating, on the other hand, it makes more sense to focus on the improvements of others. As a final step we therefore divide the improvement potential value by the current rating and thus get a ranking which shows us what we should concentrate on (see Table 4). The ranking is expressed in letters (with "A" being highest) in order not to get confused with the numbers.

| Quality component | Weighting | Feasibility | Improvement Potential (=Weighting x Feasibility) | Rating | Focus (=Potential/ Rating) | Ranking |
|------------------------|-----------|-------------|---|--------|----------------------------|---------|
| Translation quality | 5 | 3 | 15 | 4 | 3.75 | A |
| Usability | 3 | 4 | 12 | 5 | 2.4 | B |
| Translation directions | 2 | 1 | 2 | 3 | 0.66 | E |
| Supported file formats | 2 | 2 | 4 | 3 | 1.33 | D |
| Performance | 3 | 3 | 9 | 4 | 2.25 | C |

Table 4: Ranking of Improvement Focuses

7 The Road to Improvement

The ranking of improvement focuses conducted above serves as a guideline when it comes to deciding on the next steps. Apart from translation quality, it suggests that usability should be the first applicability component to address since it can easily be improved. Whether and when the other issues should be tackled must be decided separately. It goes without saying that the ranking of all components has to be reviewed regularly. Over time new ideas will enter the overall framework and the general conditions will probably change, which might deliver a totally new picture. In this regard the ranking is a good instrument to keep track of what has been achieved and also to recollect important aspects to consider.

7.1 Translation-oriented improvement

Dictionary building has always been a main pillar of establishing CLS Machine Translation. During the first two years we invested predominantly in the adding of unknown words and client-related terminology. Later we focused on more translation analyses, as described above, so that grammar and context become more involved. As the above ranking expresses there is still a lot of potential in this field. In addition to client-related terminology and the improvement of translations on a sentence level, general vocabulary is a constant "building site" too. The end user's perception of translation quality can be very arbitrary: general vocabulary or vocabulary definitely not tailored to the client's or end user's domain can be very important by serving as "flagship entries".

As we learned from the latest survey, some users would like to communicate their own observations regarding the machine's output and thus contribute to its quality. While we might not be able to incorporate all of their findings into our lexicon work, giving them the opportunity to participate will probably have a relevant impact on their quality perception since they feel more involved and no longer at the machine's mercy alone.

7.2 Application-oriented improvement

For the time being we will concentrate on ranks A to C, which means that apart from translation quality we will be looking at the following aspects:

- Usability: We will implement a feedback form and enhance the way of setting the translation directions.
- Performance: A slightly better performance might be achieved through optimized configuration of the existing system and servers. In the medium term the servers will be replaced for newer models.

The components ranking lower than C (file formats and translation directions) will not be addressed at the moment. However, after usability and performance have been improved and the ranking reviewed, the cards will have to be reshuffled.

8 Conclusion

While the discussion of MT quality usually focuses on translation quality only, it has been shown that there are further aspects to consider. Each aspect contributes to the overall quality perceived by the users and should therefore be optimized. But since the latter have different degrees of impact and involve different efforts, this requires a good basis for decision. As a service provider rather than a technology developer, CLS Communication uses a simple and effective approach to manage the quality of its MT system. It is a sound method to rate the quality at one point in time and steer its development appropriately. As it was developed for our specific purposes it should not be applied to other systems on a one-to-one basis. But the general idea can serve as a framework which can be adapted to other providers' environments.

9 References

- Bar Hillel, Y. (1964). Language and Information. Reading, MA: Addison Wesley.
- Hovy, E., King, M. & Popescu-Belis, A. (2002): An introduction to MT evaluation. In *Proceedings of the Third International Conference on Language Resources and Evaluation*. Workshop: Machine translation evaluation: human evaluators meet automated metrics (pp.1-7). Las Palmas Canary Islands: LREC
- ISO 8402 (1994), Quality Management and Quality Assurance: Vocabulary. International Standard ISO 8402:1994, Second Edition.
- Joscelyne, A & van der Meer, J (2006). Turmoil in translation. Roadmap for managing changes in the

translation industry. URL:
<http://www.translationautomation.com/member/forum/showthread.php?t=38>

Maier, E. & Clarke, A. (2001). Evaluation of Machine Translation Systems at CLS Corporate Language Services AG. In *Proceedings of MT Summit VIII* (pp 223-229). Santiago de Compostela, Spain: EAMT.

Melby, A. (1997). Should I use Machine Translation? URL: <http://www.ttt.org/theory/mt4me/index.html>: April 2007.

Schwarzl, A. (2001). The (Im)Possibilities of Machine Translation. *European University Studies Series XIV*, Vol. 381. Frankfurt am Main: Peter Lang.

Stadler, H.-U. (2005). Lexicon-coding Workflow at CLS Communication. In *Proceedings of the 10th EAMT conference "Practical applications of machine translation"* (pp. 255-261). Budapest: EAMT.