

Evaluation and Adaptation of a Specialised Language Checking Tool for Non-specialised Machine Translation and Non-expert MT Users for Multi-lingual Telecooperation

Rita Nübel

IAI
Saarbrücken, Germany
rita@iai.uni-sb.de

Abstract

Style guides or writing recommendations play an important role in the field of technical documentation production, e.g. in industrial contexts. Also, writing recommendations are used in technical contexts together with machine translation (MT) in order to circumvent the MT system's weaknesses. This paper describes the evaluation and adaptation of a language checker deployed in the project int.unity¹. In this project, both MT and a specialised language checker were adapted to the requirements of non-expert users and a non-technical domain. The language technology was integrated with the groupware platform BSCW² to support the multi-lingual communication of geographically distributed teams concerned with trade union work. The users' languages were either German or English, i.e. the users were monolingual. We chose linguatex's server version of Personal Translator 2004 MT system for the German->English translations. The language checker CLAT³ for German and English has been developed at IAI. It is used by technical authors to support the production of high-quality technical documentation. The CLAT core system was adapted and extended in order to match the new requirements imposed by both the user profile and the subsequent MT application. In this paper, the focus will be on the assessment and adaptation of style rules for German.

Introduction

The development and use of style guides, or more restricted so-called controlled languages play an important role in the field of technical documentation production, i.e. in industrial contexts. Also, writing recommendations are used together with MT in order to circumvent the hereby anticipated MT system's weaknesses. This paper describes the assessment and adaptation of the language checker CLAT originally developed to support technical authors. In the int.unity project context, the scenario for the deployment of a language checker imposed new requirements: First, the checker software is used by "non-professional" users, i.e. they are not technical writers. This means that they have no experience of producing texts according to writing guidelines. They have never used a language checking tool, and they have never used MT to support their communication. Additionally, they produce rather unrestricted texts.

Second, the checker software is used in order to support the users when writing a document for machine translation. For the evaluation and adaptation work, two important aspects thus had to be taken into account: In addition to the well-known problem of MT unpredictability, the users' reactions when confronted with CLAT messages for text revision were as well unpredictable (cf. Mitamura and Nyberg, 2001). The remainder of this paper describes the evaluation and adaptation work that was carried out to fulfil these new requirements:

In section 2 we describe the general int.unity user scenario and the architecture. The language technology and its integration with the communication platform BSCW is described in section 3. Section 4 discusses the evaluation

of the CLAT tool. Section 5 reports the adaptation and extension of the language checker, and in section 6 we summarise the results.

The int.unity User Scenario and the Architecture

The project objective was to provide an internet platform for the geographically distributed multi-lingual telecooperation. The project partners are trade unionists from the UK and from Germany. The virtual network for the participating trade unions serves the information exchange and different forms of communication of virtual project teams who during the project phase were concerned with three different thematic areas of unionist work.

The groupware platform BSCW

The BSCW shared workspace system (Appelt, 1999) has been chosen as the project's cooperation platform. The underlying concept is the definition of so-called shared workspaces. The access to the system is password protected: users register with a user name and their password. Depending on roles specifications, users have access to shared workspaces. Documents are stored on the central BSCW server and can be accessed depending on access control options.

In int.unity, BSCW enables the definition of common areas of work for the project team members who exchange documents in English and German. Apart from document exchange and storage, the users have the possibility to discuss their topics in a number of discussion forums. On a daily basis, BSCW activities are communicated to the users via email. Figure 1 illustrates a screenshot of the int.unity project workspace.

¹ The project was funded by the European Commission in the ESF-programme of Innovative Measures under Article 6 of the European Social Fund B2-1630.

² Computer Supported Cooperative Work. The BSCW platform is a product developed by OrbiTeam GmbH (Bonn, Germany.)

³ Controlled Language Authoring Technology.

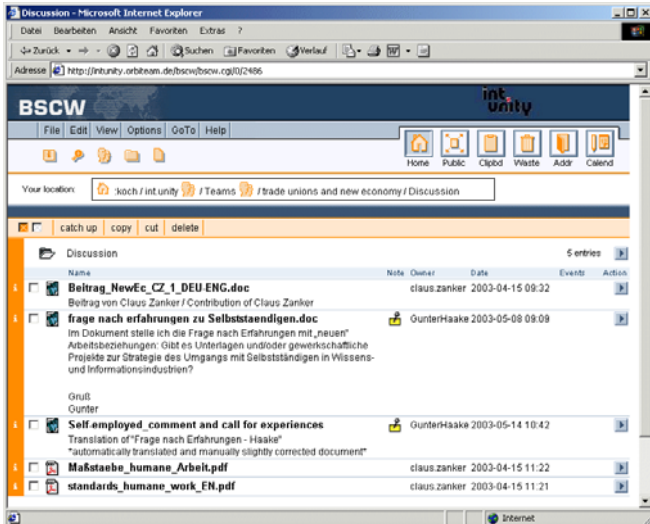


Figure 1: int.university shared workspace in BSCW.

The int.university Architecture

In int.university, the team members communicate with each other in their native languages, i.e. either German or English. Language technology (LT) is deployed in order to support the translation of the information exchange into either language. The texts to be translated should adhere to specific writing recommendations including style, terminology, grammar and orthography to yield better results of the subsequent MT. This document production is optionally supported by the CLAT system. Both MT and CLAT are accessible via the BSCW.

Language Technology in int.university

The deployed LT comprises two processes, i.e. language checking and machine translation. We have combined the commercial MT system Personal Translator 2004 together with the server software LEServe by linguattec GmbH Heidelberg, and the language checking tool CLAT which is developed at IAI Saarbruecken. The language control functions of CLAT can be activated for html-documents on the BSCW platform. There are different CLAT versions for either German or English texts. The text production, the language control and the text translation cycle in int.university comprises the following actions:

1. the user edits a text;
2. the text is uploaded in the BSCW system;
3. the user (optionally) activates the CLAT language checker via BSCW;
4. the user is notified by email from the BSCW system when the CLAT process is completed, and he can view the results of the checker in a separate html-file. This file contains a copy of the text together with CLAT annotations and messages in case there are any violations of e.g. spelling rules, grammar, terminology, or style;
5. the user revises his text according to the CLAT messages;
6. the revised text is uploaded in the BSCW;
7. the user activates the MT system via BSCW;
8. the user is notified by (BSCW)-email when the MT process is completed. The translation result is automatically stored in the user's clipboard.

Evaluation of the Language Checking Tool CLAT

The CLAT Tool

CLAT supports technical writers and editors in producing high-quality documentation. CLAT automatically checks a text and identifies spelling or grammatical errors. Additionally, term consistency and style conformity are checked, taking a corporate terminology and general as well as corporate writing rules as benchmarks. Rule violations are displayed to the user together with an error message, an example, and, in case of spelling errors and grammatical errors, of a proposal for correction. The linguistic backbone of CLAT is a morpho-syntactic analysis based on IAI's *mpro* (Maas, 1998), and a subsequent shallow parsing based on the syntactic pattern matcher *KURD* (Carl et al., 1997). Figure 2 displays the CLAT graphical user interface (GUI).

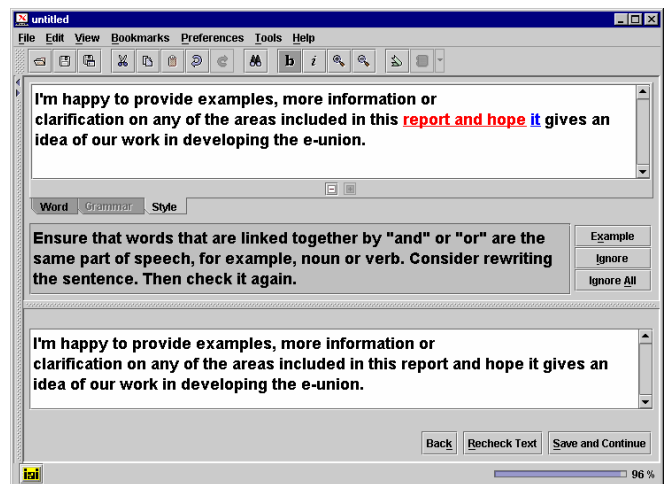


Figure 2: CLAT graphical user interface

Evaluation Criteria

New criteria for usability and functionality resulted from the user scenario and the user requirements: apart from small-scale experiments, CLAT has never been tested for its effects on MT though CLAT builds on controlled language (CL) concepts referring to intelligibility, clarity, translatability and unambiguous presentation of content (AECMA, 1998, Huijsen, 1998, Mitamura, 1999). The extension of the CLAT technology to non-technical texts and new text types was another characteristic to be accounted for. Last, CLAT is a tool for skilled technical writers, whereas in int.university the user community has no experience with text production or text revision based on CL.

These aspects lead to the following criteria for the CLAT evaluation:

- relevance or adequacy of existing CLAT specifications for MT purposes;
- adaptability and extensibility to additional MT requirements;
- relevancy or adequacy for non-technical and non-standardised texts;
- usability for non-expert users, i.e. adequacy of revision proposals.

Evaluation Strategy

Since the CLAT evaluation had to be carried out on the basis of MT results, parallel evaluation rounds were necessary for both CLAT and MT.

Test Data

The test data consisted of a corpus compiled of 21 original German texts. These texts had been produced by the users themselves. These texts belong to different genres and also show register variation, e.g.

- users' biographical information;
- informal welcome and invitations for participation
- contributions for discussion;
- short notes, questions or comments on contributions;
- protocols, agendas.

The longest text consisted of 1.159 words, whereas the shortest message contained just four words. The complete users' sample text corpus contained 7.400 words.

In addition, we chose the policy statement of the German confederation of trade unions (Deutscher Gewerkschaftsbund) as an official trade union document in order to enrich the existing corpus. This text consists of 11.350 words. For all texts, the appropriate German-English terminology was fed into the MT user dictionary.

MT evaluation

The evaluation was both sentence-based and text-based. The results of the sentence-based evaluation served the parallel CLAT assessment. The evaluation of a translated text as a whole was necessary in order to make sure that the average scores of a text translation based on sentences correlated: when a translation as a whole was acceptable, the average score calculated on the basis of the individual sentence scores should also be within the range defined for acceptable translations. For the sentence-based evaluation, translations were rated on a three-partite scale for intelligibility and meaning preservation (impressionistic rating, cf. Hovy et al. 2002):

- translation is intelligible, meaning is preserved (2)
- only parts of the translation are intelligible (2.5)
- translation is unintelligible (3)

In addition, the most disturbing errors were annotated manually in terms of an informal error description. This information served as a basis for the subsequent specification and implementation of additional CLAT rules where necessary.

CLAT evaluation

The core set of CLAT writing recommendations consists of 74 rules. The test data was run against these rules, and the rules were rated according to the criteria as described above. A high number of rules were at this stage scored *irrelevant*, since they did not apply at all. A smaller number of rules were considered as being *neutral*, because in those cases where they applied, the translation remained either acceptable or unacceptable *after* revision. A small number of rules were *inadequate* or detrimental to the translation quality. The remaining 27 rules were *relevant*,

because they had a positive effect on the MT output when the input text was revised accordingly. The table below summarises the results:

Score	Relevant	Neutral	Irrelevant	Inadequate
Nb. of CLAT rules	27	11	31	5
Phen.	complexity; sentence length; ellipsis; ambiguity; ...	layout; logical order; infinitive nominalisation; ...	compound tenses; word order; prenominal modification; ...	layout; lexicon; tense; ...
Percentage	36%	15%	42%	7%

Table 1: Relevance / Adequacy of CLAT rules.

The most frequent relevant CLAT style rules (frequency > 20) matched onto complex structures, e.g. sentence length, coordination, discontinuous verb groups, or ellipsis, but also to meta-linguistic structures, e.g. parentheses or dashes. Less frequent matches (frequency < 20) were observed with specific lexical items, ordering, expletive *es*, or modality. The least frequent rules matches (frequency < 10) covered rather specific cases of lexical ambiguity, layout, or pronouns

From the user perspective however, the most frequent matching rules were problematic because the CLAT proposals for revision remain rather unspecific. In addition, the unpredictability of MT in turn (*what makes the MT system fail?*) contributes to the unpredictability of the user's text revisions (*what does the author do?*). The following example (1) illustrates this "double black box problem":

- (1) Wo wir stehen und wie wir dahin kamen.
* Where we stand and as we got there.
Rev.1: Wo stehen wir und wie sind wir dahin gekommen?
* Where do we stand and have how we got there?
Rev 2: Wo wir stehen und: wie wir dahin kamen.
Where we stand and: how we got there.

Thus, additional, more fine-grained style rules were necessary in order to optimise the text revision process. Apart from this requirement, additional linguistic as well as meta-linguistic phenomena had to be accounted for because of the obvious. MT weaknesses in these areas.

Especially texts belonging to a more conversational register where people used a personal or familiar style required additional rules. On the other hand, more official texts contained types of structures that caused translation problems but that had not yet been accounted for.

Adaptation and Extension of CLAT

The new rules were for the most part more fine-grained in terms of structural explicitness than the existing rules.

Apart from coverage extensions, we also had to revise the matching strategy of the rules in order to allow for *overlapping* or *nested* rule application: When for a given sentence a so-called complexity rule had fired first (e.g. sentence length or coordination of main clauses, etc), the new rules then identified additional problematic structures

in this sentence, and made more explicit statements about how to revise it.

The new CLAT rules covered the following phenomena:

- relative clause constructions and ellipsis
- types of grammatical metaphor, e.g. process nominalisations in prepositional phrases (PPs) (e.g. *bei der Verwendung von X / [when] using X*)
- nested PP phrases
- layout
- specific structures occupying the vorfeld
- so-called *Füllwörter* or conversational fillers (e.g. *überhaupt, eigentlich, übrigens, nämlich*)

The strategy of nested rule application proved successful, because in many cases it reduces the user's choices for text revision and provides a better guidance for authors as exemplified below (2), (3):

- (2) Auf die Lösung dieser Probleme werden allerdings im nächsten Schritt neue Probleme folgen, die zwar heute bereits vorhanden sind, aber noch nicht offen zu Tage treten.

* On the solution of these problems will follow in the next step of new problems, though, today, although these already are existing but don't come to light openly yet.

CLAT message1: "This sentence is too long." Rule matches on the whole sentence.

Additional CLAT message2: "Avoid this type of structure at the beginning of the sentence." Rule matches on the vorfeld "Auf die Lösung dieser Probleme".

Additional CLAT message3: "Avoid ellipsis in this relative clause. Fill in the missing element(s) and split the sentence in two sentences if it is too long." Rule matches on relative clause where the relative pronoun is missing: "dabei aber nicht unkritisch sind."

After revision, the input sentence and its translation look like the following:

- (3) Wenn diese Probleme gelöst sind, wird es neue Probleme geben, die momentan nicht zu Tage treten.

If these problems are solved there will be new problems which don't come to light openly yet.

Other examples of new rules refer to e.g. conversational style / fillers (4), (5):

- (4) Aber ein Unternehmen ohne richtige Geschäftsidee zu gründen, das ist schon ganz schön gewagt.

* But set up an enterprise without a right idea of business this is already dared quite beautifully.

New CLAT message: "Avoid filler words." Rule matches on "schon ganz schön".

Revised sentence and translation:

- (5) Aber ein Unternehmen ohne richtige Geschäftsidee zu gründen, das ist wirklich sehr gewagt.

But this is really very daring to set up an enterprise without a right business concept.

Summary

In this paper we have shown the evaluation and adaptation process of the specialised language checker CLAT to a new scenario. We have shown a practical approach towards the solution of the problem of unpredictability of both MT and CLAT users: Apart from necessary coverage extensions, the improved fine-grainedness of rules considerably contributed to decreasing the users' choices for text revisions, thus providing a better guidance when rewriting a text for MT. This strategy is useful where the automatic generation of revision proposals based on explicit knowledge about the subsequent MT coverage is impossible.

References

- AECMA (1998). AECMA Simplified English, Issue 1, Rev.1. Document PSC-85-16598. A Guide for the Preparation of Aircraft Maintenance Documentation in the International Aerospace Maintenance Language. AECMA Brussels.
- Appelt, Wolfgang (1999). WWW-based collaboration with the BSCW system, In: Proceedings of *SOFSEM '99, Springer Lecture Notes in Computer Science 1725*, Milovy, Czech Republic, pp 66-78.
- Carl, Michael, Antje Schmidt-Wigger and Munpyo Hong (1997). KURD - A Formalism for Shallow Post Morphological Processing in: Proceedings of *NLPRS'97*.
- Gdaniec, C. and Arendse Bernth (2001), *Mtranslatibility*. In *Machine Translation 16 (3)*: 175-218.
- Huijsen, Willem-Olaf (1998). Controlled Language - an introduction. In *Proceedings of the Second International Workshop on Controlled Language Applications CLAW-98*, pp. 1-15, Pittsburgh (PA), USA.
- Hovy, E., Margaret King and Andrei Popescu-Belis (2002) Principles of Context-Based Machine Translation Evaluation. *Machine Translation*, 16, pp. 1-33.
- Maas, Heinz-Dieter. (1998). Multilinguale Textverarbeitung mit MPRO. In: Lobin, G. et al.(eds): *Europäische Kommunikationskybernetik heute und morgen*. KoPäd, München 1998.
- Mitamura, T. (1999). Controlled language for multilingual machine translation. In *Proceedings of Machine Translation Summit VII*, pp. 46-52, Singapore.
- Mitamura, T. and Eric Nyberg (2001) Automatic Rewriting for Controlled Language Translation. In *Proceedings of the NLPRS2001 Workshop on Automatic Paraphrasing: Theories and Applications*.

Acknowledgements

We would like to thank Thomas Koch from OrbiTeam GmbH for the provision of the BSCW illustration in figure 1. The project was funded by the European Commission in the ESF-programme of Innovative Measures under Article 6 of the European Social Fund B2-1630.