How Portable are Controlled Languages Rules?

A Comparison of Two Empirical MT Studies

Dr. Sharon O'Brien

School of Applied Language and Intercultural Studies
Dublin City University
Glasnevin, Dublin 9
Ireland
sharon.obrien@dcu.ie

Dr. Johann Roturier

Symantec Ireland
Ballycoolin Business Park
Blanchardstown, Dublin 15
Ireland
johann_roturier@symantec.com

Abstract

This paper describes two studies on the effectiveness of Controlled Language (CL) rules for MT. Both studies investigated the language pair English-German and used corpora from the IT domain. However, they differ in terms of the MT engines employed (Systran vs. IBM WebSphere) and the evaluative methodologies used. Study A examines the effectiveness of CL rules by measuring temporal, technical and post-editing effort. Study B examines the effectiveness of rules by measuring comprehensibility. Both Study A and Study B concluded that some CL rules had a high impact for MT while other rules had a moderate, low or no impact. The results are compared in order to determine what, if any, common conclusions can be drawn. Our conclusions are that rules governing misspelling, incorrect punctuation, sentences longer than 25 words, and the use of personal pronouns with no antecedent in a sentence had a high impact on both post-editing effort and comprehensibility. Further, we found that the use of personal pronouns with antecedents in the same sentence and stand-alone demonstrative pronouns had a low impact, while the rule advocating the use of "in order to" in purposive clauses had no impact in either study. The paper also discusses contrasting results for both studies.

Introduction

In the last 30 years, several initiatives have been undertaken in the field of technical communication, whereby publishers have attempted to improve the comprehensibility of their technical source content for humans or machines by implementing a Controlled Language (CL). Huijsen (1998: 2) gives the following definition of a CL: "A CL is an explicitly defined restriction of a natural language that specifies constraints on lexicon, grammar, and style."

However, deploying a large set of CL rules is sometimes difficult due to time and resource constraints. Having to check that a text conforms to a large set of CL rules takes time, even with the use of a CL checker (Govyaerts, 1996: 139). Implementing the most effective rules seems advantageous, as long as their effectiveness can be backed up empirically.

The findings of two studies conducted in this field are compared in this paper. To our knowledge, this is the first time such a comparison has been undertaken. The objective of this paper is to determine whether there is any common ground regarding the effectiveness of CL rules, in spite of differing evaluative approaches, MT systems, source texts (ST), and CL rule sets, where those rule sets did not include the explicit control of a lexicon or terminology.

Description of the two CL/MT Studies

The core objective of the first of the two studies under consideration here (henceforth "Study A") was to investigate the correlations between Controlled Language rules and machine translation post-editing effort. The justification for this research was that prior studies of the effects of CL on MT output had been restricted in scope to an evaluation of raw MT output and no consideration had been given to the effort involved in post-editing the results. Study A involved the machine translation from English into German of a User Manual for an SGML editor (1 777 source words in total). The rule-based MT system used was IBM WebSphere. The ST was written by native speakers of English and the passage selected for translation consisted of text that both described the software application and instructed the user how to perform certain functions. An example of "descriptive" text is:

ID Workbench supplies the editor mentioned as a Windows application.

An example of "instructive" text is:

1. From within Windows Explorer, double-click on a document(s).

The ST also included bulleted lists, numbered lists and frequent references to software user interface (UI) resources, e.g.:

*The **Insert Markup** dialog remains available for the user to select additional tags.*

The main objective of the second study (henceforth "Study B") was to examine the impact of specific Controlled Language (CL) rules on the comprehensibility of German machine-translated segments in order to determine whether certain rules would be more effective than others. These segments were not limited to full sentences, since they sometimes contained short series of words used in titles or bulleted lists. Whereas previous studies have focused on the

impact of CL rules on the comprehensibility of English technical documentation (Schubert et al., 1995) or on the usefulness of the resulting MT output (Bernth, 1999; Rychtyckyj, 2002), little empirical research has focused on the comprehensibility of the MT output from an end-user's perspective when no post-editing is performed. Study B evaluated the impact of 54 CL rules using a total of 304 English segments (4 463 words). These segments were extracted from consumer technical support documentation provided by Symantec, a software publisher specialised in security and availability solutions.

According to Freeman (2006: 4), technical support documentation is 'cost-reducing content' because it helps reduce customer service costs. This type of content, which is used to provide users with online solutions to technical problems, is closely related the User's Guide documentation used in Study A. In a typical technical support document, users are instructed to perform specific tasks to ensure that their application behaves as intended. One of the characteristics of this procedural style is the use of short sentences starting with imperative verbs and containing references to the UI, as shown below:

Repeat step 6 for any other file type. Click **OK***.*

Background information is also sometimes provided in a descriptive manner so that users can become more familiar with a particular topic:

A clean boot is similar to, but more thorough than, closing all applications.

In short, online technical support documentation overlaps with product documentation, but its coverage is sometimes more specific since the solutions it provides take into account precise variables (such as the interaction of a given application with third-party applications, the occurrence of specific error messages, or references to other online documents).

The results obtained in these two studies are described in full detail in O'Brien (2006) and Roturier (2006).

Methodologies

An initial assessment of overlap between eight CL rule sets for English showed that these rule sets had very few rules in common (O'Brien, 2003). It was, therefore, decided to approach Study A from a machine translatability perspective (Bernth and Gdaniec, 2001). In other words, the "machine translatability" of segments in the ST was assessed according to the existence of "negative translatability indicators" (NTIs for short) in that segment. Typical NTIs for English include passive voice, long noun phrases, ambiguous referential pronouns etc. Where NTIs occurred, the hypothesis was that those segments would require greater post-editing effort than segments where NTIs had been removed. Little research has been carried out on machine translatability assessment. The best known studies are Gdaniec (1994), Bernth (1999), Bernth and Gdaniec (2001) and Underwood and Jongejan (2001). Since Bernth and Gdaniec (2001) claim that their

approach is likely to be generalisable to different MT systems and language pairs, their list of indicators was selected as a starting point for Study A (28 NTIs were examined in total). The ST was edited in order to introduce at least two occurrences of each NTI and to reduce the occurrence of very common NTIs (e.g. pronouns). Sentences containing NTIs were interspersed with sentences containing no NTIs.

79 unknown terms were extracted from the ST and coded in the WebSphere MT system dictionary for use during the machine translation stage. The post-editing was carried out by nine professional translators, all of whom were employed by IBM and who were familiar with this text type. All post-editors had university qualifications either as translators or as linguists and the median number of years of professional experience as a translator was 14. The post-editors had little or no prior experience in post-editing MT output.

The post-editing effort was assessed, following Krings's (2001) recommendations, on three planes: temporal, technical and cognitive effort. The time taken to post-edit was captured by the keyboard logging tool, Translog (Jakobsen 1999). Technical effort, i.e. the median number of deletions, insertions, cuts and pastes per sentence type, was also measured using Translog. Cognitive effort was seen as a combination of technical and temporal effort, and the additional methodology of Choice Network Analysis (CNA) (Campbell, 2000) was used to triangulate cognitive effort results. CNA involves comparing different post-edited products for variation. Where a high level of variation occurs, that is deemed to represent a high level of cognitive effort. For a fuller description of the methodologies employed, see O'Brien 2005 and for a discussion of temporal, technical and cognitive post-editing effort, see O'Brien (2007).

In keeping with recommendations for translation process research, the post-editors were given a brief which stated that they were to post-edit the text so that:

- Any non-sensical sentences or phrases are repaired.
- Any inaccuracies in the information are fixed.
- Any mis-translation, non-translation or inconsistent translation of terminology is rectified.
- The text is understandable and stylistically acceptable to a German native speaker who needs to understand the contents of the document.

Post-editors were also informed that they were to edit the text only once and that they were not required to revise their work. While this might seem contrary to "normal" practice, it was important for the study that the correlations between NTIs and *initial* post-editing effort be captured. In other words, revision, as it is understood in the normal translation process, fell outside the scope of the study.

For Study A, a comparison of temporal and technical effort for both sentence types was carried out. In addition, the median time required for post-editing was compared with the median time required for human translation (this activity was carried out by three translators who had the same profile as the post-editors, outlined above). The latter measure is known as "Relative Post-Editing Effort" (RPE) and was put forward by Krings (2001) as a relevant comparative metric for post-editing and translation effort.

Cognitive effort was then measured for each sentence type and, more specifically, for each occurrence of an NTI using the aforementioned method of Choice Network Analysis. In effect, this meant examining the post-editing products of nine post-editors and correlating their post-editing activity with the known NTIs in a sentence. Such an analysis then allowed the researcher to group NTIs according to the criteria of "High/Moderate or Low Impact on Post-Editing". The next step involved correlating those NTIs in the "High" and "Moderate" categories with sentences that also had low processing speed (i.e. took a relatively long time to post-edit) and high relative post-editing effort measures (i.e. the post-editing effort was close to the human translation effort). This final correlation led to a list of NTIs that had the highest impact on post-editing effort. These NTIs were: gerunds, proper nouns, problematic punctuation, ungrammatical constructs, "(s)" for plural, non-finite verbs, long noun phrases, short segments, and segments that were not full syntactic units.

The classification of NTIs into High/Moderate and Low impact on post-editing will form the basis for comparison with the rules identified in Study B as having a High, Limited, or No Impact on the comprehensibility of MT output. First, however, we will describe the methodology for Study B.

In Study B, 304 segments were extracted from a corpus of technical support documentation after making sure that they violated specific CL rules. These segments were then rewritten in order to create 2 sets of segments: segments containing a CL rule violation (pre-CL segments) and reformulated segments (post-CL segments). Prior to the machine-translation stage, 197 terms were extracted and coded in the User Dictionary of a Systran WebServer 5.0 system using Systran's Intuitive Coding technology described in Senellart et al. (2001: 5). Once two sets of machine-translated segments were obtained (MT output A and MT output B), a group of evaluators was asked to read the MT output before reading the ST, and score the output by using the following criteria:

Score	Criteria
Excellent MT output (E)	Your understanding of the MT output is not improved by the reading of the ST because the MT output is satisfactory and would not need to be modified. An end-user who does not have access to the ST would be able to understand the MT output.
Good MT output (G)	Your understanding of the MT output is not improved by the reading of the ST even though the MT output contains minor grammatical mistakes. An end-user who does not have access to the ST could possibly understand the MT output.
Medium MT output (M)	Your understanding of the MT output is improved by the reading of the ST, due to significant errors in the MT output. An end-user who does not have access to the ST could only get the gist of the MT output.
Poor MT output (P)	Your understanding only derives from the reading of the ST, as you could not understand the MT output. It contained serious errors. An end-user who does not have access to the ST would not be able to understand the MT output at all.

It may be interesting to observe that a sentence can be 'comprehensible, given enough context or time to work it out,' (Coughlin, 2003: 64) by an evaluator who has access to the ST, but what about end-users who rely exclusively on the translated material? The metrics chosen in Study B therefore focused on the comprehensibility of the output from an end-user's perspective. These metrics also provided indications on the possible efforts that would be required to bring the segments to a post-edited version. For instance, by attributing an 'Excellent' score to an MT output, evaluators judged that the segment did not need to be modified. The actual post-editing task was not performed.

Four Symantec in-house translators/reviewers were selected as evaluators due to their familiarity with the topic and Symantec-specific terminology. They were asked to complete the evaluation of MT output A (304 examples for a total of 4 456 source words) before starting the evaluation of MT output B (304 examples for a total of 4 645 source words). They were not told which inputs had been treated by CL rules and they did not know whether the outputs had been mixed.

Knowing that a rule improves MT output is one thing, but knowing by how much it improves the comprehensibility of the MT output is another. In order to isolate the most effective rules, a strict scoring mechanism was designed. The evaluators' scores were replaced with numeric values. 'Excellent' was replaced by 4, 'Good' by 3, 'Medium' by 2, and 'Poor' by 1. These replacements were essential to be able to measure continuous variables, such as the median score attributed to a given segment. Any scoring agreement discrepancies had to also be taken into account to ensure that the rules' scores did not originate from inconsistent scoring. Segments with identical MT output A and MT output B were therefore excluded, and so were examples for which a consensus was not reached by at least three evaluators. An improvement had to be noted in the scores of three evaluators.

These scores were further divided to take into account the level of improvement brought by the application of a given CL rule. Two levels of positive scores were defined as follows:

For positive scores of category 1, the median score of a segment's output A had to be inferior to 3, and the median score of the corresponding output B had to be superior or equal to 3. The CL rule's rewriting ensured a major improvement in the comprehensibility of the MT output, since a majority of evaluators rated the segment's output A as 'Poor' or 'Medium' and the segment's output B as 'Good' or 'Excellent'.

For positive scores of category 2, the median score of a segment's output A had to be equal to 3, and the median score of the corresponding output B had to be equal to 4. The CL rule's rewriting ensured a minor improvement in the comprehensibility of the MT output, since a majority of evaluators rated the segment's output A as 'Good' and the segment's output B as 'Excellent'. The following formula was used to calculate the 'frequency' score of each rule:

(((Number of positive scores of category 1)*100)/Number of Segments) + (((Number of positive scores of category 2)*100)/Number of Segments)/2.

This formula confirms that scores of category 2 were not rewarded in the same manner as scores of category 1, so as to reflect the level of comprehensibility improvement brought by the CL rule. Rules were then classified as having High, Limited, or No Impact based on their frequency scores. In this study, at least 75% of the evaluators agreed on the level of improvement in 67% of examples.

Results Comparison for Both Studies

Now that we have described the different methodologies used to describe how we assessed the impact of CL rules and NTIs in the two studies under discussion, we shall address the question: What, if any, comparisons can be made across these two studies? Given that both studies involved completely different methodologies, researchers, subjects and MT systems, it is inevitable that differences will occur. Study B, for example, included 54 CL rules, while Study A included 28 NTIs. Therefore, a number of linguistic features were examined under Study B that were not examined under Study A. This led to the elimination of 25 rules which were not common to both studies. In addition, although some linguistic features were included in both studies, there were some cases where the scope of the analysis was different. For example, Study A's 'ambiguous scope in coordination' and 'multiple coordinators' NTIs were more generic than the 4 CL rules addressing coordination issues evaluated in Study B: 'Rule 22: Do not coordinate verbs or verbal phrases', 'Rule 1: Avoid ambiguous coordinations by repeating the head noun, or by changing the word order', 'Rule 25: Two parts of a conjoined sentence should be of the same type', and "Rule 23: Do not coordinate verbs or verbal phrases that share the same object when the verbs do not have the same transitivity'. Given this difference in scope, it was necessary to omit 11 features from our comparison, including the handling of –ing words. Nonetheless, this still left us with 19 features that were common to both studies and these will form the focus of our comparison here.

High Impact CL Rules

The first area where both studies had common findings was for the rule regarding misspelling. This can be problematic because MT systems cannot recognise misspelled words and usually such words remain untranslated. Study A examined 3 occurrences and Study B 6 occurrences of misspellings and the majority of occurrences had a high impact on post-editing effort and comprehensibility respectively.

Punctuation was a feature of both studies: Study A examined 17 occurrences of problematic punctuation (including incorrect use of the full-stop, colon, semi-colon, double hyphen, and the comma) while Study B looked at 11 occurrences (including the comma, semi-colon, double hyphen and question mark).

Study A found that the incorrect use of a semi-colon resulted in a high level of post-editing effort, but the other punctuation marks did not have a high impact on post-editing effort. This finding was echoed in Study B, where the semi-colon was used correctly in all pre-CL cases and the re-writing of these segments as two separate sentences did not increase comprehensibility. In Study A, the double hyphen was surrounded by spaces and the post-editing effort was consequently negligible, whereas in Study B, the double hyphen was not surrounded by spaces and this was found to have a high impact on comprehensibility due to misparsed source words, as shown below in Example 1:

Example 1:

Pre-CL ST
Type--or copy and paste--the following file names:

MT Output A
Typ--oder kopieren Sie und fügen Sie ein--die folgenden Dateinamen:

Study B examined the use of a question mark in the middle of the sentence while Study A did not. It is worth noting that using the question mark in this position leads to reduced comprehensibility in the MT output (see Example 2).

Example 2:

Pre-CL ST
In the "What do you want to call this rule?" field, type Messenger Service and click Next.

MT Output A
In dem „Wie möchten Sie diese Regel nennen??" Feld, Typ Messenger Service und klicken auf Weiter.

Post-CL ST
In the "What do you want to call this rule" field, type Messenger Service and click Next.

MT Output B
Im „Wie möchten Sie diese Regel nennen?" Feld geben Sie Messenger Service ein und klicken auf Sie Weiter.

A third rule where high impact was recorded for both studies was the rule pertaining to long sentences. Both studies examined sentences longer than 25 words (3 occurrences in Study A and B). Two out of three occurrences in Study A were shown to have a high impact on post-editing effort. The results were somewhat less clear-cut for Study B where one out of three occurrences was shown to have a high impact according to the criteria outlined under "Methodology".

While CL rules mitigate against long sentences, they also recommend that sentences should not be shorter than, for example, 4 words, as this can also be problematic for MT (Bernth and Gdaniec, 2001). Both Study A and B investigated this phenomenon and produced conflicting results (Study A found this feature to have a high impact on post-editing effort while Study B found little impact on comprehensibility). This raises an important issue with regard to the application of CL rules: the re-writing of short sentences as longer sentences can be difficult, e.g. how do you reformulate *Click OK* as a longer sentence without introducing other problems for MT?

The category of personal pronouns was divided into specific phenomena in Study B: Firstly, personal pronouns whose antecedents were not present in the segment were examined (4 cases in total); Secondly, personal pronouns referring to the preceding noun were examined and, thirdly, personal pronouns where the preceding noun was not the antecedent were examined. Where personal pronouns occurred in segments whose antecedents were not explicitly present, this was shown to have a high impact on comprehensibility. In Study A, 16 occurrences of personal pronouns were examined, only two of which correspond to the first phenomenon described above. Both of these occurrences were shown to have a high impact on post-editing effort (see Examples 3A and 3B).

Example 3A

ST
They contain everything between the two tags.

Raw MT Output
Sie enthalten alles zwischen den zwei Tags.

One example of post-edited version
Der gesamte Inhalt steht zwischen den zwei Tags.

Example 3B

Pre-CL ST
If you run LiveUpdate and it does not succeed, you will see that a link is included in the error message box.

MT Output A
Wenn Sie LiveUpdate ausführen und es nicht folgt, sehen Sie, dass ein Link im Fehlermeldungsfeld eingeschlossen wird.

Post-CL ST

If you cannot run LiveUpdate, you will see that a link is included in the error message box.

MT Output B
Wenn Sie nicht LiveUpdate ausführen können, sehen Sie, dass ein Link im Fehlermeldungsfeld eingeschlossen wird.

Low Impact Rules

While the first category of personal pronoun described above had a high impact, the two other categories of personal pronoun were found to have a limited impact in both studies. Likewise, the category of stand-alone demonstrative pronoun (7 occurrences examined in Study A, 6 in Study B). Of the 7 occurrences in Study A, only 2 were found to have an impact on post-editing effort (see Example 4) and of the 6 in Study B, none were found to have an impact on comprehensibility. Although the stand-alone demonstratives in Study B were ambiguous in the ST, this ambiguity was "correctly" transferred to the target text (TT).

Example 4

ST
This contains the title and definitions for items you will use in your document.

Raw MT Output
Dies enthält den Titel und die Definitionen für Elemente, die Sie verwenden werden, in Ihrem Dokument.

One example of post-edited version
Dieser enthält den Titel und die Definitionen für Elemente, die Sie in Ihrem Dokument verwenden werden.
Both studies examined the use of parentheses (3 examples in Study A and 6 in Study B). While Study A recorded some impact on post-editing effort, where the position of the parenthetical statement was moved by post-editors but the content remained untouched, both studies concluded that the use of parentheses had a limited impact. In fact, the rewriting of parenthetical statements may also result in a degradation in MT output through the introduction of new problems. This was observed in Study B.

Another phenomenon showing limited impact in both studies is the use of the slash as separator. 3 occurrences were used in Study A and 7 in Study B. In both studies, only 1 occurrence had an impact. In Study A, a generation problem was found when the slash separated two nouns sharing the same pronoun in the ST. This pronoun required two different gender inflections in the TT, as shown in example 5:

Example 5

ST
When you name your document/file, give it a meaningful file name.

Raw MT Output
Wenn Sie Ihr Dokument/ Datei benennen, geben Sie ihm einen sinnvollen Dateinamen.

One example of post-edited version
Wenn Sie Ihr Dokument/Ihre Datei benennen, geben Sie ihm/ihr einen sinnvollen Dateinamen.

In Study B, an analysis problem occurred when the slash separated two modifiers sharing the same head noun. In Study A, the structure *and/or* was handled properly, as was the term *OS/2*. In Study B, *and/or* was also parsed correctly, and so was the structure containing 2 nouns of different gender in the TT sharing the same definite article.

Rules with No Impact

One rule had no impact in both studies, the rule advocating the use of *in order to* to introduce purposive structures. In Study A, *in order to* was missing in 5 segments, and in Study B in 6 segments. In both studies, there was neither impact on the post-editing effort nor on the comprehensibility of the TT.

Contrasting Results

Both studies examined the impact of noun clusters, but results differed. In Study A, most clusters of three nouns or more had a high impact on the post-editing effort. Study B focused on clusters containing at least 4 nouns and found that the impact of the reformulations had little impact on the comprehensibility of the MT output. There are two explanations for this: Firstly, more noun compounds were coded in the MT user dictionary used in Study B. Secondly, prepositions were sometimes introduced in the rewritings, but failed to significantly improve the MT output.

Another rule showing different results across the two studies is the rule governing the explicit use of relative pronouns. Both studies examined at least two types of structures: when the post-modifier was either an adjective or a past-participle. In Study A, missing relative pronouns had a high impact on the post-editing effort due to the adjectival structures used in the TT, as shown in Example 6A:

Example 6A

ST
ID Workbench supplies the editor mentioned as a Windows application.

Raw MT Output
ID Workbench liefert den als Windows Anwendung enwähnten Editor.

One example of post-edited version
ID Workbench stellt den benötigten Editor als Windows-Anwendung bereit.

In Study B, however, relative pronouns were automatically inserted in MT output A by the MT system, and, therefore, no improvement in comprehensibility was registered (Example 6B).

Example 6B

Pre-CL ST
You see an error message similar to the following message.

MT Output A
Sie sehen eine Fehlermeldung, die der folgenden Meldung ähnlich ist.

Post-CL ST
You see an error message similar to the following message.

MT Output B
Sie sehen eine Fehlermeldung, die der folgenden Meldung ähnlich ist.

It should be mentioned, however, that ambiguous partitive structures were not evaluated in either study.

Study A measured post-editing effort for 5 occurrences of the passive voice. Of these, only 2 resulted in post-editing effort and in both cases the post-editing was minor and it was difficult to ascertain whether or not the effort resulted directly from the passive voice rather than from other problematic elements in the segment. Passive voice was, therefore, classified as having only a moderate impact on post-editing effort. A different conclusion was obtained in Study B, in which 12 examples were used. Three of these examples, which were rewritten using the active voice, lead to a degradation of the MT output. By removing agentive structures, part-of-speech ambiguity had been introduced in the ST.

Both studies also evaluated the impact of ungrammatical structures. Both Study A and Study B focused on lack of subject/verb agreement (3 and 2 examples respectively). Study A found that 2 out of 3 occurrences had a high impact on post-editing effort (see example 7A):

Example 7A

ST
The editor work direct with SGML files.

Raw MT Output
Die Editorenarbeit leitet mit SGML Dateien.

One example of post-edited version
Der Editor arbeitet direkt mit SGML-Dateien.

Results were not as clear-cut in Study B since one of the two violations was automatically rectified by the MT system, as shown below:

Example 7B

Pre-CL ST

WinDoctor and One Button Checkup follow strict guidelines for what they considers valid or invalid.

MT Output A
WinDoctor und One Button Checkup folgen strengen Korrekturlinien für, was sie gültig oder ungültig betrachten.

Post-CL ST
WinDoctor and One Button Checkup follow strict guidelines for what they consider valid or invalid.

MT Output B
WinDoctor und One Button Checkup folgen strengen Korrekturlinien für, was sie gültig oder ungültig betrachten.

Study B also examined the incorrect use of the possessive apostrophe, which resulted in limited impact on the comprehensibility of the MT output (one example out of two showed an increase in comprehensibility according to the evaluators).

Finally, the rule stating that *(s)* should not be used to indicate a potential plural form returned different results across the two studies. Study A found that two out of three instances had a high impact on the post-editing effort. However, the four examples evaluated in Study B did not have any impact on the comprehensibility of the MT output. Once again, the MT system automatically normalised the ST by turning *(s)* into a plural prior to the translation process. This pre-processing step ensured that any revision to the post-CL segment was redundant (see Example 8 below):

Example 8:

Pre-CL ST
The following product(s) must be uninstalled before all features in Norton SystemWorks can be installed.

MT Output A
Die folgenden Produkte müssen deinstalliert werden, bevor alle Merkmale in Norton SystemWorks installiert werden können.

Post-CL ST
The following products must be uninstalled before all features in Norton SystemWorks can be installed.

MT Output B
Die folgenden Produkte müssen deinstalliert werden, bevor alle Merkmale in Norton SystemWorks installiert werden können.

Summary

Despite differences between the approaches taken in both studies (MT system, evaluation criteria, number of evaluators and post-editors), this comparison established that a number of rules had high impact on both post-editing effort and comprehensibility. These rules include misspelling, misuse of the semi-colon, question mark in the middle of segments, use of the double hyphen when it is not surrounded by spaces, sentences that contain more than 25 words, and personal pronouns whose antecedents are not present in the same segment.

Certain phenomena only had a limited impact in both studies: personal pronouns with antecedents, standalone demonstrative pronouns, the use of parentheses, and the use of slashes as separators.
One rule had no impact in both studies: the rule advocating the use of *in order to*.

Contrasting results were, however, also obtained: Noun clusters, missing relative pronouns, passive voice, ungrammatical structures, and the use of *(s)* had different levels of impact. Nonetheless, these contrasting results were explained by (1) more terminology coding in one study (NPs); (2) automatic pre-processing by one MT engine (relative pronouns, ungrammatical input and use of (s) as a plural marker); and (3) degradation in MT output by re-writing source segments (passive voice).

Recommendations

To date, most research on CL rules has been performed in isolation. In this paper, we have attempted to bring two divergent studies together so that our combined findings might be used as a point of departure for future studies. Our comparison suggests that, for the language pair English-German and the text types User Guide/Technical Support Documentation, CL rules controlling misspelling, misuse of the question mark, semi-colon and double hyphen, long sentences and personal pronouns with no antecedents will have a high impact on MT output and should be given high priority. Although these studies used a limited number of examples, it is expected that similar results would be obtained for other language pairs in a rule-based environment since the high-impact CL rules addressed common source analysis problems.

While we have emphasised differences between the approaches taken in both studies, this final comparison shows that the evaluation criteria are complementary. On the one hand, a post-editing task involves taking comprehensibility into account, and on the other hand, as indicated in Study B, an assessment of comprehensibility should consider the post-editing effort. For future research, we would recommend a combination of these two approaches.

Keeping in mind the objective mentioned at the start of this paper, implementing as few rules as possible seems desirable to avoid impacting too much on the authoring process. As discussed in the previous section, certain pre-processing replacements can be advantageous. We therefore recommend that MT developers should provide their users with a customisable pre-processing module, where language-specific or language-independent replacements could be crafted for later use. While this recommendation applies primarily to rule-based systems, it should also prove useful with data-driven systems that have been trained with controlled data.

Bibliographical References

Bernth, A. 1999. 'Controlling Input and Output of MT for Greater Acceptance'. In Proceedings of the 21st ASLIB Conference, Translating and the Computer. London. 10 pp.

Bernth, A. and C. Gdaniec 2001. 'MTranslatability'. In Machine Translation, December 2001, Vol. 16, No. 3. Dordrecht: Kluwer Academic Publishers. pp. 175-218.

Campbell, S. 2000. 'Choice Network Analysis in Translation Research'. In M. Olohan (ed) Intercultural Faultlines: Textual and Cognitive Aspects – Research Models in Translation Studies I. Manchester: St. Jerome. Pp 29-42.

Coughlin, D. 2003. 'Correlating Automated and Human Assessments of Machine Translation Quality'. In Proceedings of MT Summit IX, New Orleans, USA. pp. 63-70.

Freeman, B. 2006. 'What Content Belongs in a CMS?' In Multilingual Computing & Technology's Content Management Getting Started Guide, April/May 2006, p. 4.

Gdaniec, C. 1994. 'The Logos Translatability Index'. In: Technology Partnerships for Crossing The Language Barrier: Proceedings of the First Conference of the Association for Machine Translation in The Americas. Columbia, Maryland. pp. 97–105.

Govyaerts, P. 1996. 'Controlled English, Curse or Blessing? A User's Perspective'. In Proceedings of the First Controlled Language Application Workshop (CLAW 1996), Centre for Computational Linguistics, Leuven, Belgium. pp. 137-142.

Huijsen, W.O. 1998. 'Controlled Language: An Introduction'. In Proceedings of the Second Controlled Language Application Workshop (CLAW 1998), Pittsburgh, Pennsylvania. pp. 1-15.

Jakobsen, A.L. 1999. 'Logging Target Text Production with Translog'. In G. Hansen (ed) Probing the Process in Translation: Methods and Results. Copenhagen Studies in Language 24. Copenhagen: Samfundslitteratur. pp. 9-20.

Krings, H.P. 2001. Repairing Texts: Empirical Investigations of Machine Translation Post-Editing Processes. Koby, G.S (ed.). Kent, Ohio: The Kent State University Press.

O'Brien, S. 2007. 'An Empirical Investigation of Temporal and Technical Post-Editing Effort'. In TIS. Vol. 2, No. 1.

O'Brien, S. 2006. Machine Translatability and Post-Editing Effort: An Empirical Study using Translog and Choice Network Analysis. Unpublished PhD thesis, Dublin City University, Ireland.

O'Brien, S. 2005. 'Methodologies for Measuring the Correlations between Post-Editing Effort and Machine Text Translatability'. In Machine Translation. Vol. 19, No. 1. pp. 37-58.

O'Brien, S. 2003. 'Controlling Controlled English: An Analysis of Several Controlled Language Rule Sets'. In Proceedings of EAMT-CLAW-03, Dublin City University, Dublin, Ireland, 15-17 May 2003. pp. 105-114.

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 thesis, Dublin City University, Ireland.

Rychtyckyj, N. 2002. 'An Assessment of Machine Translation for Vehicle Assembly Process Planning at Ford Motor Company'. In S. Richardson (ed.) Machine translation: from Research to Real Users, Proceedings of the 5th Conference of the Association for Machine Translation in the Americas, AMTA 2002, Tiburon, CA, USA, October 8-12. LNAI 2499, Springer-Verlag: Berlin Heideiberg. pp. 207-215.

Senellart, J., P. Dienes, and T. Váradi. 2001. 'New Generation Systran Translation System'. In Proceedings of MT Summit VIII, Santiago de Compostela, Spain. 6 pp.

Shubert, S.K, Spyridakis, J.H, Holmback, H.K, and Coney, M.B. 1995. 'The Comprehensibility of Simplified English in Procedures'. In Proceedings of the Professional Communication Conference:. 'Smooth sailing to the Future', IEEE International 27-29 Sept. 1995. p. 171-173.

Underwood, N.L. and B. Jongejan. 2001. 'Translatability Checker: A Tool to Help Decide Whether to Use MT'. In Proceedings of MT Summit VIII, Santiago de Compostela, Spain, 18-22 September 2001. pp. 363-368.