

CHECKING TRANSLATIONS FOR INCONSISTENCY – A TOOL FOR THE EDITOR

Magnus Merkel

NLPLAB

Department of Computer and Information Science

Linköping University

Sweden

Email: magme@ida.liu.se

Abstract

Being consistent in technical translations is difficult. Using translation memory software could be one remedy, but the effectiveness of these tools rests on the assumption that most repeated source sentences correspond isomorphically to repeated target sentences. With the help of a discrepancy tool to identify inconsistencies between source and target texts, data from translations made manually and with the aid of translation memories are presented. Apart from being used as a verification tool at the postediting phase of a translation project, the discrepancy tool will measure the relative efficiency of the use of translation memory tools, as well as identify possible shortcomings with the source text.

1. Introduction

When large quantities of technical texts are being translated manually, it is very difficult to produce consistent translations of recurrent stretches of text, such as paragraphs, sentences and phrases. This can have many different reasons, for example, several translators work on different sections of the same document simultaneously, the source text is not final and may be changed at a later stage, and it may be too time-consuming or practically impossible to identify recurrent units in the source text manually. Individual translators making up a translation team will also have individual criteria for choosing a certain translation or even choosing from a set of possible translations.¹

One suggested remedy to the problem of consistency in translation is to use tools based on translation memories. When using such systems the translators translate the text interactively with a computer tool that stores and retrieves all identical source sentences with their corresponding translations, which will guide the translator towards consistent translations. It is also possible to reuse old translations stored as translation memories of previous versions of handbooks and thereby minimizing the chances of producing variant translations of the same source segments. The quality of the translation memories that are being put to use in a translation project becomes crucial for the quality of the new target text. Translation memories can be produced either by using a translation memory-based translation tool interactively or by aligning a source text with its corresponding translation with some kind of alignment tool.

Translation memory tools are used to the best advantage when the following conditions hold:

¹ In this paper the consistency problem concerns consistent translations of sentences, not consistent terminology, although these two are related. An example of a translation checker for terminological (lexical) consistency can be found in Macklovitch (1994).

1. The source text is highly repetitive internally
2. The source text is a new version of a text that has already been translated and exists as a translation memory (i.e., external repetition).
3. Repeated sentences (or segments) of source text are in principle transferable to corresponding repeated sentences in the target text.

The first two conditions can quite easily be checked automatically, which we have done in a previous study (Merkel 1992). The third one is harder to verify, but we made an attempt to do this by confronting translators with their own translations.

We conducted an empirical investigation into translators' attitudes towards consistency and variation (Merkel 1996). This study consisted of a questionnaire involving examples of real translations of software manuals and included 50 repeated sentence types that were not translated consistently. The questionnaire was sent out to 13 translators, both in-house and freelance translators at two translation agencies where the translations had been done originally. We also asked one project leader at each company and the person in charge of translations at the customer company to complete the questionnaire. The objective of this study was to see what attitudes the people involved in a translation project had towards consistency and variation in translated texts. The results showed that there was an overwhelming preference towards consistency. The only time the translators were slightly hesitant towards being consistent was when the repeated segment occurred in totally different contexts, for example when the segment was used as a heading at one place and as a cell item at another. The implication of this is that translation memories would in principle have helped the translators in being consistent. However, another result of the study points at a problem of disagreement among translators in choosing the "right" translation. Among the two alternative translations the respondents could choose from, the respondents were only in total agreement in one out of 50 examples about which should be the correct one. Furthermore, the stated motivations for choosing a particular option varied considerably among the respondents.

In other words, what the study implies is that translators prefer consistent translations, which makes translation memory-based tools look like the desired tool. However, it also shows that translators may have difficulty in accepting a previous translation from the translation memory, which may cause frustration among the translators when they are more or less forced to accept a given suggestion. One translator commented that "the translator is reduced to somebody who presses the OK button." (The possibility to edit old entries in translation memories varies from tool to tool, but in general it is not a simple process.)

One solution to this problem is to create a special kind of translation checker that is applied after all text has been translated, at the revising stage. This translation checker would highlight all inconsistencies in the translation database and make it possible for the editor to revise the translations after the initial translation has been completed. The translators would not have to be totally restricted to previous translations, instead the consistency checking could be postponed to the revising stage. Obviously, the translation checker does not have to be used in conjunction with a translation memory tool, it could just as well be applied to a fully manual translation provided that the translation database is created with some kind of sentence alignment program. Such programs exist both as commercial software (Trados' TAlign and IBM's Visual Align) and in many variants from the academic world often originating from the algorithm presented by Gale and Church (Gale and Church 1991).

Another cause of frustration among translators is that the source text is badly edited or just of poor quality. In our questionnaire many respondents expressed this as a major obstacle for producing high quality translations. In many cases, the translators actually detect unnecessary variation in the source text and "improve" the text by making these variants consistent. Given a translation database (a

translation memory or a sentence aligned source and target text), it would be just as easy to produce a list of all “inconsistent source sentences”, which would contain all the instances where the translators have identified synonym source sentences and translated these uniformly. This would constitute useful feedback to the technical writers who produced the source text, who could use this information for the next version of the source text.

Any translation of a repetitive source text can be placed somewhere between the endpoints of a consistency continuum; between maximal *consistency* and maximal *inconsistency*. In a maximally consistent translation, all repeated source sentence types have been translated uniformly and in a maximally inconsistent translation, all source sentences have different translations. As we will see in the rest of this paper, maximally consistent translations cannot, and probably should not, be found in real translations, even if they are translated with the aid of translation memory software. There are no empirical foundations for stating any preferred percentages of degrees of consistency, so the different figures provided in the study are merely used as a starting point for measuring and discussing consistency.

2. Recurrent Source and Target Texts - Independent Analysis

If a source and a target text are analyzed independently for repetitions, the result may yield that twenty per cent of both the source and target text are repetitive on the sentence level. However, does the similarity of recurrence degrees mean that the source and target repetitions actually correspond? Let us look at a source and target text see how well the most frequent sentences correspond in numbers. Five of the most frequent sentences in the English computer manual for OS/2 are shown in Table 1²:

| Sentence | FRQ |
|--|-----|
| Click mouse button 2. | 80 |
| Open OS/2 system. | 69 |
| Open system setup. | 42 |
| Press and hold mouse button 2. | 39 |
| Select the arrow to the right of Open. | 34 |

Table 1. Five frequent source sentences of OS/2 User’s Guide

If we expect a strong isomorphic relationship between repeated sentences in the source and target text, there should be similar frequency ratios in both texts for corresponding sentences. However, none of the top ten sentence types have been consistently translated in the Swedish translation. By looking at the sentence frequency lists for both the source and the target, it is not difficult to identify the correspondences, but the differences in frequency are striking:

| Source | FRQ | Target | FRQ |
|--|-----|---------------------------------|-----|
| Click mouse button 2. | 80 | Klicka med musknapp 2. | 18 |
| Open OS/2 system. | 69 | Öppna systemprogram. | 43 |
| Open system setup. | 42 | Öppna systemkonfiguration. | 38 |
| Press and hold mouse button 2. | 39 | Håll ned musknapp 2. | 16 |
| Select the arrow to the right of Open. | 34 | Välj pilen till höger om Öppna. | 30 |

Table 2. Correspondences between repeated sentences in the source and target.

There are of course translations that are less inconsistent than the example above. Here’s an example of frequency correspondences from a Swedish translation of Microsoft’s Access User’s Guide.

² The examples are taken from the ten most frequent text sentences, excluding headings and table cell text which often only consist of one or two words.

| Source | FRQ | Target | FRQ |
|---|-----|---|-----|
| Choose OK. | 37 | Välj OK. | 39 |
| From the Toolbar shortcut menu, choose Customize. | 9 | Välj Anpassa på snabbmenyn för verktygsfält. | 9 |
| Open the report in design view. | 7 | Öppna rapporten för design. | 6 |
| Open a database in Microsoft Access or switch to the database window for the open database. | 6 | Öppna en databas i Microsoft Access eller växla till databasfönstret för aktuell databas. | 4 |
| Choose the New button. | 6 | Välj Ny. | 5 |

Table 3. Examples of repeated source sentences with corresponding translations in Access UG

The differences in frequency between source and target are considerably smaller here than in the IBM translation. It would be too tedious to do the comparisons by scrutinizing sentence lists, so we designed a simple tool for this purpose: a translation checker that focuses on detecting sentence inconsistencies.

3. The Discrepancy Tool - A Translation Checker

The input to the tool is a translation database, which consists of translation pairs of sentences that have been aligned automatically and then checked manually. The records are tab separated and also contain an field for information about the sentence mapping relation, for example 1-1, 2-1, 1-2, 2-2, etc. The discrepancy analysis is performed on the actual records of the translation database. If a record holds more than one sentence, the analysis does not split up the record, instead the whole record is treated as the unit for discrepancy analysis. Optionally, all translation pairs that are not 1-1 sentence translations could be excluded from the discrepancy analysis. The default is to analyze the translation database from source to target, that is, to find inconsistent translations of repeated source sentences, but it is also possible to analyze it in the opposite direction, from target to source, which will identify a kind of “over-standardization” on part of the translators. In the last case, the tool identifies a list of source sentences that been assigned the same translation. Apart from the listings of the actual sentence types, there is also information on some general characteristics from the translation database. The example below illustrates the types of data extracted:

The information comes from the file D:\TEXTDB\XL5\xl5.alg, and has been treated as a source -> target translation.

Source sentence types: All (types): 9957 All (instances): 12589
Consistently translated (types): 9785 Inconsistently translated (types): 172
Repeated source sentence types: All (types): 860 All (instances): 3492
Consistently translated (types): 688 Consistently translated (instances): 2336
Rtype= 8.64 repeated source types/all source types
Rinst= 27.74 repeated source instances/all source instances
ICtype= 20.00 inconsistent type translations/all repeated source types
ICinst= 33.10 inconsistent instance translations/all repeated source instances
ICtot-type= 1.73 inconsistent type translations/all source types
ICtot-inst= 9.18 inconsistent instance translations/all source instances

Figure 1. Sample of output from the discrepancy tool

The information provided concerns the number of sentence types and sentence instances, number of consistently translated sentence types and instances, number of repeated source sentence types and instances and how many of these that are consistent and inconsistent. At the end of the output above, we give percentages for various kinds of data. *Rtype* describes the ratio between all repeated sentence types and all source sentence types. *Rinst* shows the same but for instances instead of types. *ICtype* and *ICinst* describe the ratio between inconsistent sentence translations and the set of all repeated source sentences. If the translation was totally consistent, both *ICtype* and *ICinst* would be zero. And

finally, *ICtot-type/inst* gives us information on the proportion of inconsistencies in relation to all source sentences.

Let us now look at an example of a source sentence that has been translated inconsistently and therefore has been identified by the discrepancy tool:

| | Sentence | FRQ | Position & mapping relation |
|----------------|--|-----|-----------------------------|
| SOURCE: | Follow the directions in the Wizard dialog boxes. | | |
| TARGET 1: | Följ instruktionerna som visas i guidens dialogrutor. | 1 | 7886 MAP 1-1 |
| TARGET 2: | Följ instruktionerna i guidens dialogrutor. | 2 | 8183 MAP 1-1, 8262 MAP 1-1 |
| TARGET 3: | Följ anvisningarna i dialogrutorna. | 1 | 10924 MAP 1-1 |
| TARGET 4: | Följ instruktionerna i dialogrutorna som visas i guiden. | 1 | 15051 MAP 1-1 |

Table 4. Repeated source sentence with four different translations

Here the sentence “Follow the directions in the Wizard dialog boxes.” occurs four times in the source text, but has been translated in four different ways, where TARGET 2 indicates that one of the alternatives has been used twice. If we take the same translation database and reverse the analysis, that is, from target to source, we get data of potential synonym source sentences. Two examples from such an analysis are shown in Table 5:

| | Sentence | FRQ | Position & mapping relation |
|----------------|---|-----|-----------------------------|
| TARGET: | Mer information finns i avsnittet "Ange relationer mellan tabeller" i kapitel 7, "Grunder för tabeller". | | |
| SOURCE 1: | For more information, see "Setting Relationships Between Tables" in Chapter 7, "Table Basics." | 1 | 695 MAP 1-1 |
| SOURCE 2: | For details, see "Creating Relationships Between Tables" in Chapter 7, "Table Basics." | 1 | 5330 MAP 1-1 |
| SOURCE 3: | For more details, see "Setting Relationships Between Tables" in Chapter 7, "Table Basics." | 1 | 8943 MAP 1-1 |
| TARGET: | Välj Kör på Fråga-menyn eller klicka på Kör i verktygsfältet. | | |
| SOURCE 1: | Choose Run from the Query menu, or click the Run button on the toolbar. | 1 | 6180 MAP 1-1 |
| SOURCE 2: | From the Query menu, choose Run (or click the Run button on the toolbar). | 1 | 6232 MAP 1-1 |
| SOURCE 3: | Choose Run from the Query menu (or click the Run button on the toolbar). | 2 | 6416 MAP 1-1, 6493 MAP 1-1 |

Table 5. Two target sentence types with three synonym sources each

4. A Study of Translation Databases

We have run the translation checker on eight different texts, two novels, two manually translated software manuals, three software manuals translated with the aid of a translation memory tool, and one MT-translated collection of dialogue fragments. The two novels are not really interesting applications for consistency checking, partly because they are not at all as repetitious as the manuals and partly because consistency is not something necessarily aimed for in literary translation.

Nevertheless, the data from the novels could prove to be interesting for translation studies. The MT-translated dialogues are not included here as they show no traces of repetition at all, and consequently there are no inconsistencies to be measured. Instead we will concentrate on the software manuals. Let us begin with looking at the data from the manually translated software manuals, Excel User's Guide and Access User's Guide from Microsoft:

| CATEGORY | Access | | Excel | |
|-------------------------------|--------|-------|-------|-------|
| | S->T | T->S | S->T | T->S |
| Sentence types | 10849 | 10970 | 9957 | 9950 |
| Sentence instances | 14704 | 14704 | 12589 | 12589 |
| Consistent types | 10502 | 10771 | 9783 | 9586 |
| Inconsistent types | 347 | 199 | 174 | 194 |
| Repeated types | 1272 | 1111 | 860 | 846 |
| Repeated instances | 5127 | 4845 | 3492 | 3685 |
| Consistent repeated types | 925 | 912 | 686 | 652 |
| Consistent repeated instances | 3384 | 3290 | 2332 | 2357 |
| Rtype | 11.72 | 10.13 | 8.64 | 8.68 |
| Rinst | 34.87 | 32.95 | 27.74 | 29.27 |
| ICtype | 27.29 | 17.91 | 20.23 | 22.93 |
| ICinst | 34.00 | 32.09 | 33.22 | 36.04 |
| ICtot-type | 3.20 | 1.81 | 1,75 | 1,99 |
| ICtot-inst | 11.85 | 10.58 | 9.21 | 10.55 |

Table 6. Discrepancy data for two manually translated software manuals

The data in Table 6 show the discrepancies in both directions, source to target, and target to source, indicated in the columns S->T, T->S respectively. The texts are roughly of the same size, Access being slightly more repetitious than Excel. When we look at the consistency measures, we find that the Access translation shows more signs of being inconsistent (*ICtype*: 27.29 vs. 20.23 per cent). The Excel translators have also identified more inconsistencies in the source than the Access translators (*ICtype*: 22.93 vs. 17.91), but these are only small differences. What is clear is that both translation teams have "missed" a certain number of sentences that could have been translated consistently, but they have also "improved" the translation in relation to the source text by making synonymous source sentences have the same translation.

The Microsoft manuals are far from being 100 per cent consistent; if that had been the case, we would have had zero values for *ICtype*, *ICinst*, *ICtot-type* and *ICtot-inst*, both from source to target, and from target to source. For manual translations, however, this is what one could expect. Several translators working simultaneously on different parts of the documents must face problems to be consistent if they do not have the support of a translation database. What is remarkable, in spite of the lack of computerized support, is that the figures for inconsistencies from target to source are as high as they are, which means that the translators somehow have identified a large set of different source sentences that have been translated uniformly.

Let us now turn to two translations that have been made with the aid of translation memories and compare them with the manual translations. A reasonable expectation here would be that the figure for *ICtype/inst* would drop to close to zero, approaching maximal consistency when applied from source to target, and that there will be high figures for *ICtype/inst* when applied from target to source. This last expectation depends highly on the use and success of the fuzzy matching techniques present in the translation memory tool. With fuzzy matching the translators ought to have good opportunities to detect minor differences in the source text and thus make the translation more uniform. The translations were made with the aid of IBM's Translation Manager/2 and stem from the period when this tool was introduced into the translation process at IBM Sweden.

| CATEGORY | OS/2 IG | | OS/2 UG | |
|-------------------------------|---------|-------|---------|-------|
| | S->T | T->S | S->T | T->S |
| Sentence types | 2615 | 2640 | 7054 | 7148 |
| Sentence instances | 3057 | 3057 | 8876 | 8876 |
| Consistent types | 2521 | 2567 | 6774 | 6914 |
| Inconsistent types | 95 | 73 | 280 | 234 |
| Repeated types | 190 | 196 | 685 | 638 |
| Repeated instances | 631 | 613 | 2507 | 2366 |
| Consistent repeated types | 95 | 123 | 405 | 404 |
| Consistent repeated instances | 252 | 324 | 1086 | 1240 |
| Rtype | 7.26 | 7.42 | 9.71 | 8.93 |
| Rinst | 20.64 | 20.05 | 28.24 | 26.66 |
| ICtype | 50.00 | 37.56 | 40.88 | 36.68 |
| ICinst | 60.06 | 47.15 | 56.68 | 47.59 |
| ICtot-type | 6.63 | 2.77 | 3.97 | 3.27 |
| ICtot-inst | 12.40 | 9.45 | 16.01 | 12.69 |

Table 7. Discrepancy data for OS/2 Installation Guide and OS/2 User's Guide (TM tool)

The two texts are considerably shorter than the Microsoft manuals, and they are slightly less repetitious if we measure the percentage of repeated sentences. Nevertheless, there are some strange things going on here. If we look at the *ICtype* figure (measuring the number of inconsistently translated sentences in relation to the set of repeated sentences), we see that the OS/2 translations are much more inconsistent than the Microsoft translations, in spite of the use of Translation Manager/2. In the *OS/2 Installation Guide*, fifty percent of all repetitions show signs of inconsistencies. In the *User's Guide* the corresponding figure is lower, but still higher than in the Microsoft translations, which were made without a translation memory tool. The *ICtype* figures in the reverse direction indicate however that the translators have found many inconsistencies in the source text and standardized this variation in the translation. The data in Table 7 seems to indicate that the use of the translation tool had very little effect on the translation as far as increasing consistency from source to target text. Furthermore, it raises questions on whether translation memory software really saves time and money, as is often promised by the producers of such software (cf. Schäler (1994)).

From the data alone we could not figure out what had really happened during the translation project so we had to go back to the translators at IBM. In an interview with two of the translators who had been working with the company's translations before and after the introduction of TM/2, the following explanations were given:

IBM had in the beginning of the nineties had a long history of "rewriting". The source text was merely seen as some kind of guidance of what the translator/writer should produce. The goal was a coherent Swedish text, clearer and more concise than the original. What they aimed for was something that in Newmark's terminology was something of a "free translation" with a strong target language emphasis (Newmark 1988). In the translation guidelines for IBM, it is stated that the translator should review the source text critically, delete unnecessary passages or repetitions, rearrange and regroup the source text before starting the translation (Ström and Windfeldt 1991). The translation culture at IBM in the beginning of the nineties was not exactly suitable for a swift adaptation to translation memory-based software. They had a translation team that was highly skilled in producing high-quality Swedish versions of American originals in a creative way. Some translators had an openly negative attitude towards the new tools, some people left and some protested in other ways, by for example ignoring TM/2 and sticking to the old ways of doing things. Initially there were also problems with the administration and distribution of translation memories, which also contributed to resistance from the translation teams.

The basic explanation to the strange discrepancy figures in the OS/2 translations according to my interviewees was that there had been a clash between new technology and an old translation culture, where the old culture initially was stronger than new tools.

Instead we looked at some other texts where they thought that the use of TM/2 had been more successful. These texts are IBM's documentation on InfoWindows and Client Access for Windows, where the translation of the former actually stems from about the same time as the OS/2 translations, but the latter was done during 1995. A summary of the discrepancy data for these texts looks like the following:

| CATEGORY | InfoWindows | | Client Access | |
|-------------------------------|-------------|-------|---------------|-------|
| | S->T | T->S | S->T | T->S |
| Sentence types | 5157 | 4881 | 1689 | 1658 |
| Sentence instances | 7771 | 7771 | 2054 | 2054 |
| Consistent types | 5121 | 4443 | 1663 | 1617 |
| Inconsistent types | 136 | 438 | 26 | 41 |
| Repeated types | 1076 | 1214 | 146 | 153 |
| Repeated instances | 3590 | 4104 | 882 | 920 |
| Consistent repeated types | 940 | 776 | 120 | 112 |
| Consistent repeated instances | 2974 | 2213 | 718 | 687 |
| Rtype | 20.47 | 24.87 | 8.64 | 9.26 |
| Rinst | 46.20 | 52.81 | 36.37 | 37.94 |
| ICtype | 12.64 | 36.08 | 17.81 | 26.80 |
| ICinst | 17.16 | 46.08 | 18.59 | 25.33 |
| ICtot-type | 2.59 | 8.97 | 1.54 | 2.47 |
| ICtot-inst | 7.93 | 24.33 | 6.76 | 9.61 |

Table 8. Discrepancy data for InfoWindows and Client Access

Here the degree of inconsistencies is much lower from source to target, 11.64 and 17.81 per cent (*ICtype*), compared with the other IBM translations. The translators working for IBM said during the interview that they had seen significant changes in IBM's translation culture in Sweden. Now there was more pressure to reuse old translations, and adjust the translation to this recycling process. One example of this is that the more recent translation of Client Access contains 98.35 per cent 1-1 mappings (one source sentence - one target sentence) whereas for the "older" OS/2 Installation Guide the corresponding figure was 81.91 per cent. This in itself is a good indication of the changes that have taken place.

5. Other Applications

Restricting variation in the source text is also used for MT systems that adopt a controlled language approach. If a company moves in the direction of automatic translation, a discrepancy analysis of their previous translations will give useful information on how to design such a controlled language. Consider for example the variation of the source sentences in Table 5.

With the increasing availability of text corpora, researchers, teachers and students within corpus linguistics and translation studies have access to new ways of exploring theoretical and descriptive branches of their fields, see for example Baker (1993). One direct application of the discrepancy tool would be to study variation of equivalent idiomatic dialogue in different languages, based on translation corpora. As mentioned earlier, the novels included in the analysis do not contain a high degree of repeated sentences. But the small set of repetitive sentences that do occur is interesting, because, at least in one of the two novels, the repetitions are from the fictitious characters' dialogue, which give an interesting contrast between idiomatic dialogue patterns in English and Swedish. If large translation corpora containing dialogue were created, a discrepancy analysis would provide a

starting point for studying idiomatic correspondences which go far beyond the phraseology, style and conventions described in lexicons and translation text books.

6. Conclusions

Translation memory tools are designed for translation projects of repetitive texts and should in principle help the translators to increase both speed and quality of the translation task. However, if these tools will prove to be efficient depends on the quality of the source text, the quality of previous translation memories, the attitude towards new technology among the translators using such tools as well as to what extent repeated source text segments can be transferred to corresponding target segments.

In this study we have shown that both manually translated texts and texts translated with the aid of translation memories are more or less inconsistent. For manually translated texts the variations are what can be expected, but the high degree of inconsistency found in IBM's early translation memory translated manuals was due to a clash between an established translation culture and new technology. Over time the use of translation memories seems to have been more successful, if efficiency is measured as the degree of consistency.

The discrepancy tool should be seen as a translation checker that can be activated in the postediting phase of a translation project. The tool will help the editor to pinpoint inconsistencies in the translation, edit them or leave them as they are in the published translation. The editor will furthermore have the possibility to verify the translations in a translation memory before it is archived for future use or remove inconsistencies in an existing bank of many translation memories. This is important when the number of translation memories is steadily growing. If translation memories are not verified at one stage or the other, the translator may be facing so many translation alternatives that browsing through these options will take as much time as translating the sentence without the memory.

The discrepancy tool could also be seen as a complement to translation memory software and one of several utilities that will enhance the quality of the translations. Ideally, discrepancy analysis should be integrated with the translation memory software, in order to verify and clean up the memory before archiving. It will help to make the translations more consistent, and also to extract information about unnecessary variation in the source text which can be fed back to the technical writers. Applying the tool to translation databases will give a clear indication of how efficient the use of translation memory tools has been. If the inconsistencies between source and target are too numerous, then this may reveal that there is a serious problem in the company's management and configuration of the translation project.

Acknowledgments

This work is part of the on-going project "Linguistic Engineering for Generation and Translation of Documentation" at Linköping University. The project is financed by the Language Technology Program jointly managed by the Swedish Council for Research in the Humanities and Social Sciences (HSFR) and The Swedish National Board for Industrial and Technological Development (NUTEK). I would like to thank Microsoft WPGI in Dublin and IBM Sweden for providing us with their translation material. Also, I am grateful to Lars Ahrenberg for his comments on previous drafts, to the three anonymous reviewers of this paper for valuable suggestions for improvement. My thanks also extend to Sara Nordberg who did the first implementation of the discrepancy tool, and to Mikael Holm and Mathias Olsson who improved the tool by giving it a graphical interface.

References

- Baker, M. 1993. Corpus Linguistics and Translation Studies – Implications and Applications. In *Text and Technology*, edited by M. Baker, G. Francis and T.-B. E. Philadelphia/Amsterdam: John Benjamins Publishing Company.
- Gale, W., and K. W. Church. 1991. A program for aligning sentences in bilingual corpora. In *Proceedings of the 29th Annual Meeting of the Association for Computational Linguistics*.
- Macklovitch, E. 1994. Using Bi-textual Alignment for Translation Validation: the TransCheck System. In *Proceedings from AMTA-94*.
- Merkel, M. 1992. Recurrent Patterns in Technical Documentation: Department of Computer and Information Science, Linköping University.
- Merkel, M. 1996. Consistency and variation in technical translations - a study of translators' attitudes. Paper to be published in the *Proceedings of Unity in Diversity, Translation Studies Conference*, Dublin.
- Newmark, P. 1988. *A Textbook of Translation*. London: Prentice Hall.
- Schäler, R. 1994. A Practical Evaluation of an Integrated Translation Tool during a Large Scale Localisation Project. In *Proceedings of ANLP-94*. Stuttgart.
- Ström, A.-M., and K. Windfeldt. 1991. *Klartext på NLSG*: IBM Svenska AB.