

# SYSTRAN Review Manager

**Jean-Cédric Costa**

SYSTRAN Software Inc.

9333 Genesee Avenue, Suite PL1  
San Diego, California, USA  
costa@systransoft.com

**Christiane Panissod**

SYSTRAN S.A.

1, rue du Cimetière  
95230 Soisy-sous-Montmorency, France  
panissod@systransoft.com

## Abstract

The SYSTRAN Review Manager (SRM) is one of the components that comprise the SYSTRAN Linguistics Platform (SLP), a comprehensive enterprise solution for managing MT customization and localization projects. The SRM is a productivity tool used for the review, quality assessment and maintenance of linguistic resources combined with a SYSTRAN solution. The SRM is used in-house by SYSTRAN's development team and is also licensed to corporate customers as it addresses leading linguistic challenges, such as terminology and homographs, which makes it a key component of the QA process.

Extremely flexible, the SRM adapts to localization and MT customization projects from small to large-scale. Its Web-based interface and multi-user architecture enable a centralized and efficient work environment for local and geographically disbursed individual users and teams. Users segment a given corpus to fluidly review and evaluate translations, as well as identify the typology of errors. Corpus metrics, terminology extraction and detailed reporting capabilities facilitate prioritizing tasks, resulting in immediate focus on those issues that significantly impact MT quality. Data and statistics are tracked throughout the customization process and are always available for regression tests and overall project management. This environment is highly conducive to increased productivity and efficient QA in the MT customization effort.

## 1 Introduction

Automation of the translation process is the only viable solution for most applications with high volumes of dynamic content, such as technical support. Yet many companies resist using MT because of quality concerns and the challenge for determining return on investment, as the measurement for MT quality is subjective and always depends on differentiating variables. Often, human translators oppose using MT for a variety of reasons. But the quality issue is almost always central to the debate.

The SYSTRAN Review Manager (SRM) is a key component of SYSTRAN's customization methodology. It provides a centralized framework for building and maintaining customized linguistic resources in a consistent and structured manner. It also guarantees that quality assurance is an integral part of the MT customization process.

SYSTRAN actively engages with corporate customers to ensure they understand all phases of the MT workflow. This is critical as most organizations have little, if any prior experience dealing with MT and therefore are unable to effectively foresee challenges. The SRM inadvertently forces corporate users to become involved in the evaluation of translation quality and the entire MT process.

The SRM addresses the following issues:

1. Evaluation of MT output.
2. Improvement of MT output.
3. Measurements for translation quality – good or acceptable translation at the post-editing stage.

## 2 Review interface

SRM's core feature is the review and evaluation of a translated corpus – segmented into Translation Units (TUs) – at different stages of the MT customization process.

Reviewers work on automatically generated HTML reviews forms that display:

- The *source TU* and frequency in the corpus.
- The *translated TU*, in which unknown words (NFWs), customized dictionary entries, and other information are easily identifiable by color codes.
- The *review criteria*: defined by the administrator and used by the reviewer. The criteria combine options from drop-down lists and/or free-form text fields.

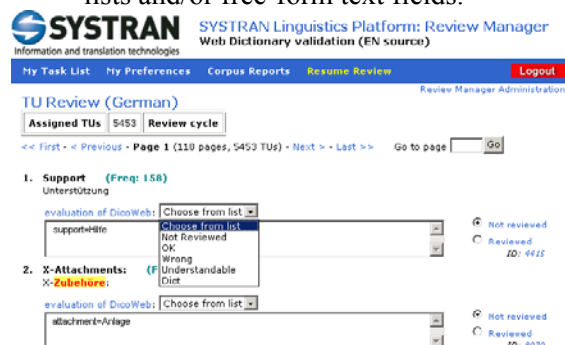


Figure 1: Translation review task

Additional information on each TU is available to the reviewer by clicking on the TU unique identifier. This information includes metrics (words, characters, frequency, etc.), list of documents in which the TU appears, and the TU context.

### 3 Review typology

There are two main types of translation review tasks:

- After an MT system update: used to check for regression and measure the new translation quality level.
- For MT system maintenance: used to further improve translation quality by translation review cycles and the addition of new dictionary entries.

The review typology is fully customizable. This enables the SRM to be extremely flexible, suitable for both linguists and non-linguists, and adaptable to a variety of review requirements.

#### 3.1 Quantitative assessment: rating

The simplest form of a review criterion is boolean: “Correct translation” or “Incorrect translation”. This is suited for fast, yet efficient user ratings with bilingual skill-sets and knowledge of the

terminological domain (linguistic expertise is not required). For this typology, the average review speed is about 150 TUs per hour.

However, the review criterion is often refined to include a typology of error for “Incorrect translations”. As the bulk of the customization work is terminology related, the item “Incorrect translation” is often replaced by the following values: “Requires a Do Not Translate (DNT) entry” or “Requires a multilingual entry”. A further detailed error typology (such as morphology, homograph, word order, etc.) may be included for reviews performed by linguists.

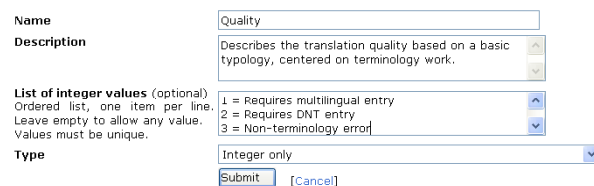


Figure 2: Simple review criterion

The SYSTRAN Translation Quality Metric is used when the customization work focuses on technical support texts. This metric is based on the Society of Automotive Engineers (SAE) J2450 translation quality metric, used to measure human translation but adapted for MT use. A weighted numerical score represents the average translation quality of a TU, based on the number of errors in 3 categories: grammar, terminology and format. With such a typology, the average review speed drops down to about 80 TUs per hour.

#### 3.2 Qualitative assessment

Review criteria may include a text field that allows users to provide qualitative feedback on the translations, such as free-form comments. However, such feedback usually requires manual analysis by a linguist.

#### 3.3 Dictionary enrichment

Qualitative feedback about the translation results is used to improve the translation quality. The SRM takes advantage of SYSTRAN’s IntuitiveCoding® technology, able to turn a review criterion’s text field into a Bilingual Dictionary Editor. While rating the translation, reviewers may add new terminology using this field. The average speed for this type of review is about 50 TUs per hour.

For example, an English to German entry:  
support (noun) = Hilfe (noun) (feminine)

and some Do Not Translate (DNT) entries:  
 Oracle (company name)  
 Mike (proper noun) (masculine)

### Reports and Statistics

Review Cycle:  and target language:

- Source corpus profile
- Translated corpus profile
- NFW list [View] [Download]
- Review status report
- Review data extraction
- User Dictionary generated by the reviewers [View] [Download] [Compile]

**Figure 3:** Reports and statistics page for UD compilation

The bilingual list of words, compounds, or expressions introduced by the reviewers is exported to the SYSTRAN Dictionary Manager (SDM) – another component of the SYSTRAN Linguistics Platform (SLP) – for further processing. Alternatively, it is automatically compiled into a runtime User Dictionary (UD).

The resulting UD is then used to re-translate the source corpus. New translations are quickly compared to previously translated TUs. The new UD terminology is displayed on-screen in green. A Microsoft Word-like revision mode emphasizes the changes.

4. STP might put one of the links in the blocking mode to avoid loops . (Freq: 1)  
 Le STP pourrait mettre une des liaisons dans le mode de blocage pour éviter des boucles.  
 Le STP pourrait mettre une des liaisons dans le mode *blee de blocage* pour éviter des boucles.

Terminology:  Grammar:  Format:

**Figure 4:** Checking for improvements and regressions

The reviewer validates that no regressions were caused by the UD entries and that the translation results meet user expectations.

## 4 Corpus

A *review project* is associated to an MT customization effort, characterized by a specific source language and a corresponding source corpus. The *corpus* is indexed, analyzed by the SRM and segmented into unique Translation Units (TUs). TU frequency and many other metrics are computed and used to generate a corpus profile.

As MT customization projects often require work on large corpora, only a statistically significant part of the corpus is reviewed. It is especially important to be able to focus on those issues that bear the biggest impact on translation quality.

### Source Corpus Profile

The corpus contains 30 documents.

|         | TUs   | Unique TUs | %       | Words  | Characters | Size    | Text ratio |
|---------|-------|------------|---------|--------|------------|---------|------------|
| Minimum | 20    | 20         | 100.00% | 237    | 1243       | 1292    | 96.21%     |
| Maximum | 10935 | 10835      | 99.09%  | 162711 | 901437     | 956971  | 94.20%     |
| Average | 1600  | 1544       | 96.50%  | 22884  | 126714     | 150918  | 83.96%     |
| Total   | 48006 | 45947      | 95.71%  | 686524 | 3801425    | 4527538 | 83.96%     |

There are 45947 unique TUs out of 48006 total TUs (i.e. a 95.71% unicity).

|         | Frequency | Words  | Characters |
|---------|-----------|--------|------------|
| Minimum | 1         | 1      | 1          |
| Maximum | 155       | 120    | 839        |
| Average | 1.04      | 15     | 82         |
| Total   | NA        | 678218 | 3765799    |

**Figure 5:** Source corpus profile (partial)

The SRM offers an advanced corpus subset extraction tool that allows users to:

- Extract high frequency sentences.
- Extract sentences with NFWs.
- Extract sentences with a specific word, phrase or regular expression.
- Extract only sentences with a change in the translation, observed when comparing two states of the MT system.

#### New TU Subset

Please define your query using the fields below. The selection criteria to be ignored should be left empty. For numerical fields: 12 for an exact match, 1-7 for a range, <56 for a maximum and >4 for a minimum. Once you are satisfied with the results you can save the subset by checking the corresponding checkbox and choosing a subset name.

Optional: select a review cycle to have some selection criteria that apply to the translated TUs.  
 Review cycle:  and target language: .  
 The previous cycle is: SYSTRAN CSD only translation baseline #2 - 2003/07/21.

**Criteria on source TU**

Words:   
 Frequency:   
 Size:   
 Contains:   
 Does NOT contain in order to:

**Criteria on previous translated TU**

NFW:   
 Dict. entries:   
 Dict. coverage:   
 Contains:   
 Does NOT contain:

Search:

**Criteria on translated TU**

NFW:   
 Dict. entries:   
 Dict. coverage:   
 Contains:   
 Does NOT contain:

**Translation comparison criteria**

NFW:   
 Dict. entries:   
 Dict. coverage:   
 Same translation:

**Figure 6:** Corpus subset extraction

## 5 Review cycles

The review process is a succession of separate and independent *review cycles*. A review cycle is associated to a specific state of the MT system development. The typology and target languages of each review cycle may vary depending on review needs.

During corpus translation, Not Found Words (NFWs) are automatically extracted along with their frequency ratios. A translated corpus profile is also generated. It includes NFW distribution and dictionary coverage. The NFW list may be exported to the SDM for coding.

Translations and metrics for all review cycles are accumulated in the relational database that underpins the SRM. They are always available for regression tests and project management purposes.

## 6 Review tasks and reports

Once a review cycle is defined and the corpus is translated, the review administrator creates and assigns review tasks to the reviewers, i.e. a specific subset of the corpus.

### New Task

Review cycle SYSTRAN CSD+UD translation comparison #2 - 2003/07/21  
 Target language Spanish  
 Task name Test review task  
 Reviewer X example  
 Task  TU review of the TU subset ES comp (3184)  TU review of the document subset -- Choose from list --  
 Subset range 1-1000 (document range for document subset, TU range for TU subset)  
 Use the format 1-5 to assign the first 5 TUs of the subset; nothing means "all".

Figure 7: Task creation

Several reports allow the administrator to efficiently supervise the MT customization effort.

**Review Status Report**  
**Task Summary**

| Cycle                     | Target  | Tasks | Reviewers | TUs * | Unique TUs * | First review               | Last review                |
|---------------------------|---------|-------|-----------|-------|--------------|----------------------------|----------------------------|
| Cisco evaluation training | Spanish | 10    | 3         | 881   | 881          | 2003-07-21 16:45:36.640252 | 2003-07-23 16:33:41.629266 |

\* Number of TUs assigned to reviewers

**Review Summary**

| Reviewed TUs (unique) | Total TU reviews | % of tasks | % of corpus |
|-----------------------|------------------|------------|-------------|
| 44                    | 44               | 4.32 %     | 5.00 %      |

**Score**

| Minimum | Average | Maximum | Std. dev. | Variance |
|---------|---------|---------|-----------|----------|
| 0.00    | 0.26    | 1.03    | 0.07      | 0.26     |

**Review Detail (per criterion)**

**Format**

| Reviewers | Unique TUs | Min. | Avg. | Max. | Total | > 0       | Std. dev. | Variance |
|-----------|------------|------|------|------|-------|-----------|-----------|----------|
| 44        | 44         | 0    | 0.02 | 1    | 1     | 1 (2.3 %) | 0.15      | 0.02     |

**Grammar**

| Reviewers | Unique TUs | Min. | Avg. | Max. | Total | > 0         | Std. dev. | Variance |
|-----------|------------|------|------|------|-------|-------------|-----------|----------|
| 44        | 44         | 0    | 0.41 | 2    | 18    | 15 (34.1 %) | 0.62      | 0.39     |

**Terminology**

| Reviewers | Unique TUs | Min. | Avg. | Max. | Total | > 0         | Std. dev. | Variance |
|-----------|------------|------|------|------|-------|-------------|-----------|----------|
| 44        | 44         | 0    | 0.91 | 4    | 40    | 23 (56.8 %) | 1.03      | 1.06     |

**Review Detail (per reviewer)**

| Username | Tasks # | TUs # | Reviewed TUs | %       | Avg. score | Min. score | Max. score | Score std. dev. | Score variance | First review               | Last review                |
|----------|---------|-------|--------------|---------|------------|------------|------------|-----------------|----------------|----------------------------|----------------------------|
| edgeres  | 1       | 100   | 19           | 19.00 % | 0.26       | 0.00       | 0.84       | 0.18            | 0.03           | 2003-07-22 17:00:01.707339 | 2003-07-22 17:28:13.931516 |
| netesene | 1       | 100   | 0            | 0.00 %  | 0          | 0          | 0          | 0               | 0              | NA                         | NA                         |

Figure 8: Review status report

Once a reviewer is logged into the SRM, his/her task list is automatically displayed with a task completion percentage and other task characteristics. By clicking on the task name, the review forms are displayed (see figure 1). Several display preferences are available, such as side-by-side, line-by-line, the number of TUs per page, or other options.

## 7 Conclusion

The SRM empowers SYSTRAN customers as it provides them with the means to build and maintain their own linguistic resources. The SRM

is also used internally by SYSTRAN for small to large-scale MT customization projects, as well as new language pair development, like the Arabic, Swedish and Danish systems.

Driven by customer and internal needs, the development of new releases for the SRM is ongoing. With the advent of next-generation XML-based SYSTRAN engines and the continuous improvement of SYSTRAN's IntuitiveCoding® technology, the opportunities expand: generalization of source and target markup – currently limited to NFWs and dictionary entries – support for new formats and new language pairs.

1. حالة من الطيات الضخام العارم من أقاليم صيدية المظفرة بعد حوالي 50 كلم غربا العاصمة العراقية بغداد بعد أن سافطت عليهم 15 قذيفة خلال الأسبوع الماضي  
 The situation from the explosion the debris the massacre changed to ... protect on lengthened the road against the occupying the American, at when says rudder the city indeed some relatives the victims stood by burial their without received consolation waits for the retaliation.

2. البصرة لاجرا بقتلهم دون تعيل العرا الضحايا لا اذ النار  
 The victims from the massacre from martyrs the massacre changed to ... protect on lengthened the road against the occupying the American, at when says rudder the city indeed some relatives the victims stood by burial their without received consolation waits for the retaliation.

3. الضحايا ومن شاركهم من ردها العيشة انكروا أن صيدية أقاليم كانت مطقة المتعاقبة برحيل القوات الأمريكية التي دخلت هنا يومين إلى الضحية  
 The victims and those who participated in their death the massacre denied that the American forces entered the city since days to the city.

Figure 8: Using the SRM for the development of the Arabic to English system

## 8 Bibliographical References

Richardson, Stephen, William Dolan, Arul Mezenes, Jessie Pinkham. 2001. *Achieving commercial-quality translation with example-based methods*. MT Summit VIII, Santiago de Compostela, Spain. Pp293-298.

Rychtyckyj, Nestor, 2002. *An Assessment of Machine Translation for Vehicle Assembly Process Planning at Ford Motor Company*. AMTA Conference, Tiburon, CA, USA

Senellart, J., Yang, J., Rebollo, A., 2003. *SYSTRAN Intuitive Coding Technology*. MT Summit IX, New Orleans, Louisiana, USA.

Senellart, J., Boitet, C., Romary, L., 2003. *XML Machine Translation*. MT Summit IX, New Orleans, Louisiana, USA.

Underwood, Nancy L., Bart Jongejan. 1999. *Profiling Translation Projects*. TMI-99, Chester, England. Pp139-149.