Translation Quality Control at the FBI

Focus on Feedback and Intelligence Value

Elizabeth Rhoades, MA, J. Isaiah Harbison, PhD, Sarah C. Wayland, PhD, Allison L. Powell, PhD, Barbara Forsyth, PhD

Translations serve an essential function at the FBI. They serve not only to aid agents in understanding and creating a criminal case, but also as important evidence in the subsequent trial (Fishman, 2006). Having an accurate translation, an accurate representation of what was said in the source language, is crucial.

Starting in 2008, the University of Maryland Center for Advanced Study of Language (CASL) conducted a study to determine whether providing a human-generated quick transcript in the same language as the audio would allow FBI linguists to create verbatim translations more quickly and/or with greater accuracy (Wayland, et al., 2009). The study was commissioned to help the FBI decide how to assign transcription and translation work tasks and to aid in staffing decisions.

To maximize the relevance of that study to the FBI, we designed our translation scoring methodology to work with their existing Quality Control (QC) process. This document describes the FBI’s method, and provides a brief overview of other translation quality assessment methods, showing the principal strengths of the FBI’s approach, specifically:

- The emphasis on and identification of errors that impact the intelligence value of the translation, and
- The identification and correction of errors for the purpose of mentoring.

We hope that this document will serve as a useful resource for the FBI and any individual or agency interested in learning about the FBI’s QC process.

FBI TRANSLATION QUALITY CONTROL

The Quality Control (QC) Review process developed by the FBI is a method used to assess and monitor the accuracy of translations, to improve translations, and to mentor linguists. The primary purpose of the FBI’s QC Review process is to make sure the translation preserves the intelligence value (meaning) of the original transmission.

Quality Control Review Methodology

Each translation that goes through the QC process is assigned to a certified QC Reviewer. To become certified, linguists attend and successfully complete the rigorous FBI Quality Control Review training. The QC Review training emphasizes the preservation of intelligence value, or meaning. A translation that conveys the same information as the source language material is satisfactory. QC Reviewers are trained to differentiate between equivalencies and errors and to assess the severity of any identified errors. Even a translation with errors could be considered satisfactory, as long as the errors do not have a catastrophic impact on intelligence.
Translation Equivalencies

Equivalencies are expressions that contain different words or different word order but provide the same information; for example,

(1) The meeting took place in the library.
(2) The library was the location of the meeting.

Although the sentences are organized differently and contain different words, they convey the same information and would be considered equivalent. QC reviewers are trained to accept equivalencies—even if a QC Reviewer believes that sentence (2) would be a closer translation of the source language material, if sentence (1) provides the same information, it should not be marked as an error. Two words that have the same meaning would also be considered equivalencies; however, QC Reviewers are encouraged to check a dictionary if there is any doubt whether one word is really equivalent to another.

Quality Control Error Notation Key

QC Reviewers are also trained on the FBI’s Quality Control Error Notation Key (QC-ENK), the guide used for classifying errors (see Figure 1, below).

![Quality Control Error Notation Key (QC-ENK)](image)

**Figure 1.** The FBI’s Quality Control Error Notation Key (QC-ENK)

There are only two major error categories, accuracy errors and expression errors. Accuracy errors involve specific information such as proper nouns, numbers, dates, and technical terms. For example, a wrong date is considered a Key Mistranslation and given the code KMT. Expression errors are errors in grammar, punctuation, capitalization, or spelling. The QC-ENK guide presents each error code with detailed examples so that all accuracy and expression errors in a translation can be coded consistently. Only the types of errors categorized on the QC-ENK are marked on a translation. All other errors, such as those regarding genre, register, or formatting, are not marked, but are addressed later in the QC Review process, in the reviewer’s feedback.
The Quality Control Review Process

The three major steps to QC Review process are explained in detail in the following sections.

Step 1: Identify, label, and correct errors

First, a QC Reviewer carefully reads the translation and identifies and accepts all equivalencies. Then the reviewer uses the Quality Control Error Notation Key (QC-ENK) to identify, label, and correct the accuracy and expression errors. Each error is underlined, given an error code, and corrected in erasable red ink. Figure 2 shows an example of how errors are marked and corrected.

![Figure 2](example.png)

Step 2: Review Errors and Compose Feedback

After the errors are marked and corrected, the QC Reviewer composes the feedback that accompanies the translation as a part of the QC Review. The objectives of the feedback are to discuss patterned errors and errors that catastrophically affect intelligence, to mentor the linguist and improve the translation, and to determine a rating for the translation. The feedback section is therefore not simply a list of errors, but also includes a discussion of the errors in the translation.

Step 2a: Identify Patterned Errors

Before writing the feedback, the QC Reviewer goes through the translation, paying close attention to the errors that were marked, to see if there are any patterns that emerge. If there are any error patterns, the QC Reviewer describes them, explains why the errors are errors, and explains how to fix them. When relevant, the QC Reviewer also explains why and how the errors affect the intelligence value of the translation. The following feedback excerpt from a QC Review describes a pattern of errors and the impact on intelligence value:

“There are patterns of errors in accuracy in the form of omissions and numerous changes to the text. These errors affect the intelligence value of the product since they involve the drug transaction which is the topic of the conversation. For example, the original text mentions that the drug is cooked a lot in that place. The translation states, "I think that money is cooked here." These numerous errors have a catastrophic impact on the intelligence value of the translation since they provide incorrect and inaccurate information on the content of the conversation.”

Step 2b: Identify Catastrophic Errors

After identifying any patterned errors in the translation, the QC Reviewer goes through the translation to see if there are any isolated errors that could have a catastrophic impact on the intelligence value of the translation. An error is considered to have a catastrophic impact on intelligence if it results in a loss of meaning that is not recoverable elsewhere in the translation. Therefore, in order to determine if an error has a catastrophic impact on the intelligence value of the translation, the QC Reviewer must go through the translation to see if the meaning
that is lost as a result of the error is recoverable elsewhere in the translation. If there are any catastrophic errors, 
the QC Reviewer explains why the errors are errors based on the QC-ENK, why and how the errors affect the 
intelligence value of the translation, and how to fix them.

Any type of error could have a catastrophic impact on intelligence, even an error that might otherwise appear 
to be minor. For example, if the word “lima” was in a translation when it should have been “Lima,” this 
capitalization error could have a catastrophic impact on the intelligence of the translation because, unless the 
meaning is recoverable from other parts of the translation, the reader would think that the translation was about a 
type of bean instead of the capital of Peru. This example illustrates why the primary concern of the FBI QC 
Review process is not the necessarily the number or types of errors, but the preservation of meaning for 
intelligence value. Below is an example of feedback describing an error that catastrophically affects the 
intelligence of a translation:

“This translation contains accuracy errors (omissions, changes) that affect the intelligence value of the 
product. For example, the sentence in the foreign language material “he is coming around three-thirty in the 
afternoon,” which states when the man will come with the merchandise, is not present in the translation. This 
 omission has a catastrophic impact on the intelligence value of the translation since there is therefore no 
information as to when the transaction will take place.”

**Step 2c: Compose Feedback**

The feedback is composed of three sections: a summary of the accuracy errors, a summary of the expression 
errors, and a section that describes any errors not on the QC-ENK. In the accuracy and expression error sections, 
the QC Reviewer explains why the errors are considered errors based on the QC-ENK, why and how the errors 
affect intelligence, and how to fix the errors. In the first two sections, the QC Reviewer also discusses patterns of 
accuracy and expression errors, which could collectively affect intelligence.

The QC Reviewer uses the third section of the feedback to discuss any errors not listed on the QC-ENK. 
These include errors regarding conventions of the target language, errors in register, genre, and style or 
formatting errors cited in the Manual for Standards of Translation (MST). The MST is an FBI publication that 
outlines all of the standards of translation and interpretation established by the FBI, such as formatting and style 
guidelines. Most of the errors discussed in this section would not have an impact on intelligence, although 
sufficiently high numbers could render the translation not satisfactory.

The objective of the feedback section of the QC Review is not only to improve the translation, but to mentor 
the linguist who created it. One way the QC Review accomplishes this is by requiring QC Reviewers to provide 
an explanation of how to fix the errors. In addition, the QC Review feedback is written in active voice. The 
feedback does not address the linguist directly, but refers instead to the errors contained in the translation. For 
example, the phrase “This translation contains capitalization errors” would be used instead of “The first letter of 
every sentence was not capitalized.”

The feedback section of the QC Review is also used to determine the overall rating of the translation. The 
feedback must therefore present and summarize the translation errors in a way that indisputably justifies the final 
rating.

**Step 3: Rate the translation**

After composing the feedback, the QC reviewer rates the translation. Each translation that goes through the QC 
Review process is given a rating of “satisfactory” or “not satisfactory.” The rating is based on the number of errors 
and whether or not the errors affect the intelligence value of the translation. The rating should follow logically from 
the feedback. A single error in the translation that has a catastrophic impact on the intelligence value of the translation 
would render the translation “not satisfactory.” If there are a number of errors that together have a catastrophic impact 
on intelligence, then the translation is rated “not satisfactory.” A translation may also be rated “not satisfactory” if 
there are numerous translation errors, even if the errors do not have a catastrophic impact on intelligence. Figure 3 is 
a passage from a translation that was rated “not satisfactory” due to the large number of errors.
ALTERNATIVE TRANSLATION QUALITY ASSESSMENT METHODS

Groups and organizations other than the FBI are interested in translation quality, but their approaches tend to have different goals. Translation Quality Assessment (TQA), the measurement of how well a translation captures the original text, has become a topic of research in its own right (Arango-Keeth & Koby, 2003). However, while most translators work with semi-technical and technical texts (Hammond, 1990), much of the research in TQA has focused on the translation of religious, philosophical, and political texts (Williams, 2001). Emphasis on these sorts of texts has had an impact on the assessment measures discussed in the TQA research literature.

The predominant method of formal translation assessment consists of error marking (Arango-Keeth & Koby, 2003). Error marking procedures identify and label each error according to a specific classification system. For example, Thelen (2006) presents an error marking procedure with seventeen types of errors, ranging from specific errors in grammar and word order to more general errors such as style errors or rephrasing into “translationese.”

Error marking methods can also provide a system for marking the severity of an error, called error weighting. This is a quantitative extension of error marking where the weight assigned to each error indicates the severity of the error. While there are several methods of assigning weights, most weight the error based on some sort of classification system. For example, a “wrong term” error in a translation might count as three points against the translation, whereas a misspelling might count as only one point against it.

Instead of having a single, fixed value for errors of each type, some methods specify a range of potential weights for each type of error depending on its impact on the translation. An example of this method is the SAE J 2450, a translation quality metric used by the automotive industry, which allows one value to be assigned to an error if it is serious and a lesser value if it is minor. Within such a system, the potential weight of an error is still heavily dependent on its classification. The maximum weight given to a punctuation error could only count as much as the least significant minor syntactic error.

For translation quality assessment purposes of the FBI, these error counting and error classification methods miss something critical, a means of measuring the accuracy of a translation relative to intelligence value. From this perspective, even a minor punctuation error could change the meaning—and catastrophically compromise the intelligence value of a translation. Such an error must be weighted heavily.

---

1 The concept of ‘translationese’ is defined as ‘deviance in translated texts induced by the source language’ (Johansson & Hofland 1994, 26).

For example, the use of the “she-cat” in the following translation might be perceived as translationese:

Translation: The pupils of the she-cat grew large in the darkness.
Source text: La gata dilataba las pupilas en la oscuridad. (Hervey, Higgins & Haywood, 1995).
The American Translators Association uses an error-weighting process. In addition to classifying 22 error types, it allows an evaluator to assign an error weight of 1, 2, 4, 8, or 16 (Secara, 2005). However, this is still a relatively fixed system that does not fully gauge the impact of an error on translation meaning.

Others in the translation community have recognized the importance of meaning maintenance. Williams (2001) described a meaning-based method of assessing translations of arguments by comparing a translation to the original source material in terms of how many of the critical parts of the argument were maintained. Overall, however, there had been no method developed in the translation community that met the unique intelligence-impact requirements of the FBI.

CONCLUSION

The Quality Control Review Process developed and used by the FBI uses an approach that separates it from all other TQA methods: the impact of an error is measured solely by its impact on intelligence value. The final rating is determined by how much the errors compromise intelligence. In addition, feedback and rating provisions create a format for ongoing mentoring, quality control and translator improvement.

REFERENCES


Acknowledgment: We wish to thank May Kamalick for teaching us the particulars of the FBI’s Quality Control Review process. We would also like to thank Shana Semler and Adriana Peroutka for sharing their knowledge of the FBI’s QC process with us. And last but not least, we thank Margaret Gulotta for her support throughout all phases of this project.