BOLT Program: Egyptian Arabic to English Translation Guidelines

Version 1

September 12, 2013

Linguistic Data Consortium

NOTE: If for any reason translators are uncomfortable working with any particular document included in their assignment, please contact LDC at translations@ldc.upenn.edu to request a replacement file.

1 Introduction

These guidelines provide an overview of requirements for translation of Egyptian Arabic text into English to support BOLT Phase 2 evaluation of machine translation technology. Genres evaluated in BOLT Phase 2areSMS (text messages) and chat.. LDC collects data from donations of SMS/chat archives and live-messaging through our collection platform.and selects certain conversations for translation.

This document describes the format of the source text as well as its translation and addresses specific issues when translating discussion forum text.

2 Translation Teams

Initial translation is performed by translation services employing professional translators. Services must assign files to be translated to a team consisting of at least two members: an Arabic dominant bilingual and an English dominant bilingual. One team member does the initial translation, while the other one proofreads the translation. It's up to the translation agencies to decide who does the initial translation and who does the proofreading. The team makeup may not change during translation of a particular data set.

A translation service may have multiple teams working simultaneously. Proofreaders may be shared among teams unless LDC provides instructions to the contrary for a particular project.

Translation teams may use an automatic machine translation system and/or a translation memory system to assist them during translation.

Translation teams must be documented as follows:

- Translator and proofreader profiles consisting of name or pseudonym, native language, second languages, age and years of translation experience. When multiple translation teams are used, also indicate team membership for each person.

- Work assignment information consisting of the team number or the name of the translator and proofreader for each file in the data set.

- The name and version number of any translation system or translation memory used.

- A description of any additional quality control procedures or other relevant parameters or factors that affect the translation.

This documentation should be submitted to LDC along with the completed translations.

3   File Formats

The source text for translation comes in many different data formats and may include such metadata as speaker labels, timestamps,message boundaries or other information. LDC converts all source text into a standard translation format before sending data out to translators in order to 1) make the source files easy for translators to read; 2) avoid translators' tampering with the metadata; and 3) aid automatic processing of the data after the translation is returned to LDC.

3.1   Source file

Each source file delivered from LDC is divided into segments that roughly correspond to sentences or sentence-like units. Each Arabic segment in the source file consists of 2 components:

> **<ar=##>** is a unique identifier for the Arabic segment;
> **Arabic text to be translated**.

Each corresponding English segment in the source file consists of one component:

> **<en=##>** is a unique identifier indicating where the English translation should be added for a given Arabic segment.

The ## for a given English segment is identical to the corresponding Arabic segment to be translated.

A sample source file appears below:

```
<ar=10>  يا فندم يـعني إيـه شرعية؟؟ هو الـميدان طلب يـحط قـوانـيـن وبـعيـن وزرا
!!ويـحط الـدستور؟؟
<en=10>
<ar=11>  الـميدان والـشارع عايـز يـسقط حكم الـعسكر لـابـد وحالا.. الـبـرلـمـان عنده
!!!اعتـراض علي كده؟؟
<en=11>
```

3.2 Completed translation file

The completed translation should be formatted exactly the same as the source file; the only difference between the source file and the completed translation file is that an English translation will be added after each English segment identifier.

Translators should type the English translation after each "<en=##>" tag <u>without altering any other part of the file</u>. Altering anything in the Arabic segment or adding unnecessary line breaks, carriage returns or other stray marks in the files makes it difficult for LDC to automatically post-process the translation files. Similarly, in cases where a single Arabic sentence is translated into multiple English sentences, no blank lines should be inserted between the English sentences.

To summarize: the content inside square brackets should not be translated or copied over into the translation. Other content should be translated.

English translations should be rendered in plain ASCII text using UTF-8 encoding.

A sample of a completed translation in the correct format follows:

```
<ar=23> انا سالت جوزى قال انه احتمال يكون عنده نوع من الجنون
<en=23> I asked my husband. He said that he might be suffering from some
kind of insanity
```

3.3 File Naming Conventions

It is very important that completed translation files use the exact same name as the initial file provided by LDC. Please do not add anything to the file name (like ENG or the agency name) or change anything in the file name, including the file extension.

3.4 Packaging deliveries for LDC

Part 1: Before sending LDC translated files, please put them into a folder that has the same name as the package we sent you. For example:

```
If we send you a package named
bolt_phase1_arabic_set4_NNN.zip [NNN being the name of your agency]

please name the folder containing the translated files
bolt_phase1_arabic_set4_NNN
```

Part 2: Then zip the folder and make sure the ZIP file has the same name. For example:

```
bolt_phase1_arabic_set4_NNN.zip
```

4   Delivery of Completed Translations

Completed translation files should be submitted electronically, as zipped email attachments, via FTP or by web upload as specified by LDC. Paper transmission is not acceptable. All completed translation files must be in plain text, not in some proprietary software format (like Microsoft Word).

Translations must be delivered on time according to the schedule negotiated with LDC at the start of the project. Late submissions may result in payment penalties. If it looks like a translation delivery may need to be later than the agreed-upon submission date, please contact LDC as soon as possible to discuss the situation. In some cases, a partial delivery may be required before the final delivery date.

4.1   Email Correspondence about Translation

When sending correspondence about a translation project, please include the following information in the subject header:

- Agency Name
- Translation Package Name and Part Number (where applicable)
- A descriptive phrase like "Translation Delivery", "Invoice", "Question" etc.

For instance, the following subject headers are helpful:

```
ABCAgency BOLTTrain01.Arabic Part 1 – Translation Delivery
AgencyX BOLTPhase2Eval.Arabic Part 2 – Invoice
```

Using a descriptive header helps LDC direct your message or delivery to the right place, ensuring timely response.

5   Specific Rules for Translation

5.1   General Principles

The goal of the BOLT Phase 2 Translation process is to take Arabic source text drawn from SMS/chat conversations and translate it into fluent English while preserving all of the meaning present in the original Arabic text. Translation agencies will use their own best practices to produce high quality translations. While we trust that each agency has its own mechanism of quality control, we provide the following specific guidelines so that all translations are guided by some common principles:

- The English translation must be faithful to the original Arabic text in terms of both meaning and style. The translation should mirror the original meaning as much as possible while preserving grammaticality, fluency and naturalness.

- Try to maintain the same speaking style or register as the source. For example, if the source is polite, the translation should maintain the same level of politeness. If the source is rude, excited or angry, the translation should convey the same tone. SMS/chat data is largely informal and this register should be maintained in the translation.

- The translation should contain the exact meaning conveyed in the source text, and should neither add nor delete information. For instance, if the original text uses *Bush* to refer to the current US President, the translation should **not** be rendered as *President Bush, George W. Bush,* etc. No bracketed words, phrases or other annotation should be added to the translation as an explanation or aid to understanding.

- The translation should also respect the cultural assumptions of the original. For example, if the Arabic text uses the phrase *Hosni Mubarak*, the translation should **not** be rendered as *President Hosni Mubarak*—instead, it should keep the term used in the original source.

- All files should be spell checked and reviewed for typographical or formatting errors before submission.

5.2   Proper Names

Proper names should be translated using conventional practices.

Whenever an Arabic proper name has an existing conventional translation in English, that translation should be used. For example, *Gamal Abdel Nasser* (جمال عبد الناصر), the late former president of Egypt, should be translated as *Gamal Abdel Nasser*, not *Jamal Abdel Nasser* as Modern Arabic would have suggested.

The order of first name and last name presentation for the name in the source file should be preserved. For instance, if the source file reads *Osama Bin Laden*, this should **not** be changed to *Bin Laden Osama* in the translation.

Non-Arabic proper names should be translated as they would be translated into English directly from the original language. In case of an original English name appearing in the Arabic text, the normal English form should be used.

Recall that speaker IDs that appear inside square brackets are provided to aid comprehension and should **not** be translated or copied over into the translation.

Occasionally, the translator may observe that the name appearing within the speaker ID tag is spelled incorrectly. Translators should always use the correct English spelling/transliteration of a name within a file, even when the speaker ID for that name is incorrect.

Lacking preexisting knowledge of how to translate a proper name, the translator should consult a standard resource or do a quick web search. This is especially important for names that re-occur, for instance, names that are part of a news story or names of political leaders. For names of "regular people" that occur only once *and* are highly unlikely to be found on the web (for instance, the names of the author of a newsgroup post), translators should use their best judgment about how to translate the name, following standard conventions. Lacking information to the contrary, translators should simply proceed as if the name were an Arabic name.

For specific proper names, such as names of agencies, programs, conferences, books, films and other media, translators should follow the generally accepted or most commonly used form. If no common form exists or if there are multiple forms in frequent use, translators should provide the translation that follows linguistic rules instead of a translation that is word-for-word but incorrect or awkward.

Names should be translated consistently within and across files.

5.3   Numbers

As a general rule of thumb, numbers in the translation should appear either spelled out in full, or written as ASCII numbers, according to how they appear in the source text. However, there are some general conventions for writing numbers in American English that should be followed.

In American English, commas are generally used for numbers with more than three digits unless the number represents a specific year (for instance, the year 1984). Following is an example of when a comma should be used:

<div dir="rtl">احنا مصنعناش التاريخ لمدة طويلة يمكن توصل لألف وربعميه سنة.</div>

```
We have not made history for a long period of time that goes back 1,400
years.
```

Use a combination of numerals and words for very large numbers:

سعر البيت ده ١٥٠٠٠٠٠,٠٠ دولار.
```
This house costs 1.5 million dollars.
```

5.4   Capitalization

Translators should follow standard written English rules for capitalization unless there is strong evidence in the source text that suggests a different treatment. Proper names should be capitalized, including personal names and names of organizations and geo-political entities. The first word of each sentence should also be capitalized.

انت <mark>بتقول</mark> شيء و الواقع له رأي مختلف تماما فالغالبية التي هي الشارع اختارت اعضاء مجلس الشعب
```
You say something and the reality is something else. The majority,
which is the street, chose the members of the Parliament.
```

طبعا انا اتلبشت لكن فوجئت ان موظف الشباك بيقول مهو **بنك مصر** لسة مسروق امبارح كمان!!!!!!!!!
```
Of course, I was confused but I was also surprised to hear the cashier
say Masr Bank also was robbed yesterday!!!!!!!!!
```

5.5   Punctuation

Written standards for punctuation vary across languages. As a general rule of thumb, punctuation in the translation should follow standard English punctuation conventions. Punctuation in the source text primarily serves to enhance readability, so translators should not spend too much time worrying about the exact placement of commas and internal punctuation in the English translation.

The BOLT program will tackle unstructured genres, so non-standard punctuation is expected to prevail. SMS and chat pose a lot of challenges in terms of punctuation. For these two genres, translators should **not** standardize punctuation in the translation when it is missing from the source text. Non-standard punctuation (dashes, ellipses, etc.) should also be preserved in the translation.

Similarly, in SMS/chat data there will be instances of non-standard punctuation and missing punctuation. Generally speaking, non-standard punctuation in the source should be preserved in the target. However, sentences at the end of an SU may be

handled differently than sentences elsewhere in the SU. The following describes how translators should approach missing end-of-sentence punctuation at the end of an SU.

NOTE: In some cases, a single Arabic SU may result in an English SU containing multiple sentences. The following applies to the end of the SU:

1. IF the sentence is <u>at the end of the SU (i.e. the final or only sentence of the SU)</u> and has no end-of-sentence punctuation,

   THEN:
   - If the sentence is a question, add a question mark.
   - If the sentence is **not** a question, **don't** add any end-of-sentence punctuation.

2. IF the sentence is **not** <u>at the end of the SU (i.e. sentences that precede the final sentence)</u>,

   THEN:
   - Provide standard end-of-sentence punctuation (for example, in the case that an Arabic SU has resulted in multiple English sentences).

5.6 Idioms

Idioms, colloquial expressions and the like are particularly difficult to translate for human translators, let alone for machine translation engines. To help machine learning, we will provide both intended meaning and literal meaning of the idiomatic expressions. If a similar expression exists in English, you should use it as the intended meaning. The literal meaning should always be rendered in fluent English, rather than as a word-for-word translation. The translations should be surrounded by [ ] as in [text1 | text2] with | separating the two translations. Text1 should be the intended meaning and text2 should be the literal meaning. For example:

```
Format:     [ intended meaning | literal meaning ]

NOTE: Intended meaning should always precede literal meaning.
```

إيش جاب لجاب جمال عبد الناصر لحسني مبارك
```
[There is no such comparison | what did he bring to bring] between
Gamal Abdel Nasser and Hosni Mubarak
```

هو ده زي القطة بسبع أرواح يعني يوقع من البلكونة وما يمتش

```
He is like [a cat has nine lives | like a cat has seven lives], I mean,
he fell down from the balcony but didn't die.
```

لو تبص على مزيعة قناة النيل هي زي قمر اربعتاشر

```
If you look at Nile Channel's =anchor, she is [beautiful like a full
moon | like the moon in the 14th night]
```

أخوية متعلم وخريج جامعة بس سبع صنايع والبخت ضايع. بقالو سنة يدور على شغل ومش لاقي

```
My brother is educated and has a college degree-- [Jack of all trades,
master of none | seven trades and no luck]. He has been looking for a
job for a year and hasn't found anything.
```

انا بكلمك عن الثورة المصرية وانت بترجعلي للتاريخ. طب ما اللي فات مات وإحنا اولاد النهارده.

```
I am talking to you about the Egyptian Revolution and you keep going
back to history. [Let bygones be bygones | what has passed is now dead]
and we are the sons of today.
```

5.7   English, or other language content

Occasionally, English, other dialects of Arabic or another language may appear in the source text.

English sentences in the source text should be copied over to the English translation exactly as they appear in the source text. <u>Do not make any changes or corrections to the English</u> even if the English contains grammatical or other errors. Non-English foreign text needs to be copied exactly in the translation without any change. There is no need to translate such texts in English. For example:

```
<ar=1> He have a good handwriting.
<en=1> He have a good handwriting.
```

Dialectal Arabic words (other than Egyptian dialect) in the source text should be translated with the intended meaning. If translators are not sure of the real meaning of the dialectal words, they should make their best guess about the appropriate translation and surround the translation with (( )) to indicate the best guess. Refer to 5.10 for the use of (( )).

5.8 Factual Errors in Source Text

Factual errors in the source text should be translated as is. They should **not** be corrected. This also applies to grammatical errors or other speaker "mistakes" in the source text. For example:

<ar=1> يعني رئيس السلطة المصرية محمود عباس مش حيقدر يواجه حماس
<en=1> Mahmoud Abbas, the President of the Egyptian Authority, is not going to be able to confront Hamas.

<ar=4> يعني اول ما استلم حسني مبارك رئاسة الجمهورية المصرية بالستينات اقتصاد البلد كان بخير
<en=4> When Hosni Mubarak took over presidency of the Egyptian Republic in the 1960s, the economy was doing okay.

5.9 Typographical errors and misspellings

Translators will occasionally notice obvious typographical errors or obvious incorrect use of homophones in the source text. In such cases, translators should translate the intended meaning but should <u>add the flag = before the translated word to indicate that it is a correction of a typo.</u> For example:

<ar=1> المعنة اللى بتديها الصين لافريقيا
<en=1> the =aid that China provides to Africa

Be careful to distinguish obvious typographical errors, which should be corrected, from factual errors in the source text, which should **not** be corrected. If it is not clear whether the item is a typographical error or a factual mistake, translators should **not** correct the item.

Because there is no standardized spelling in Egyptian Arabic, it may be challenging to determine when there is a misspelling. As a general guideline, translators should rely on context to help them determine whether a given word is misspelled. The following are examples of misspelled words:

المفرود يكون فى انتماء علشان الناس تحس اننا بنحب بلدنا وان كل اللي بيتقال فى التليفزيون والمسلسلات كذب بدليل اننا بنحب بلدنا

There =should be a sense of belonging so that people feel that we love our country and that all what is said on TV and in soap operas are lies and the proof is that we love our country.

أن يجمع الجميع على شخص على شخص بعينه فهذا هو المستحيل بعينه ولكن خلينا مع اللي بتتفق الأغلبية على أنه أفضل من غيرة

For everyone to agree on one person, this is inconceivable, however,

```
let us be with the one who gets the agreement of the majority that he
is better than =the others.
```

5.10 Difficult to Translate Source Text

In rare cases, the source text may be so difficult to understand that translation is very difficult. In such cases, translators should make their best guess about the appropriate translation, but should surround the translated text with (( )) to indicate that this is a guess based on confusing source text. For instance:

<ar=1> الإنتخابات العربية ما بتجريش على مرشح واحد بيفوز بميه في الميه من الأصوات زى ما فاز القائد
الضرورة قبل ما يسقط

```
<en=1> Arab elections are not held with one candidate who gets a
hundred per cent of the vote, just as the ((bloody)) leader did before
his fall.
```

<ar=5> شحتفونا لما خدوا البقرة من البيت، واللي لقوه اخدوه.

```
<en=1> ((We grieved)) when they took the cow from home. They took
everything they found.
```

As always, translators should use available resources, including the internet, to find the most appropriate translation for unfamiliar terms or phrases.

5.11 Special Issues for Translation of SMS/chat

Translating SMS/chat conversations may prove challenging, as many factors may influence the linguistic style, terminology, grammar and punctuation used in the source text , and there is potential for great variation within even a single post.

The challenge for translators when dealing with SMS and chat messages as source texts is to be aware of the different linguistic styles within the conversation. Furthermore, it is important that translators acknowledge any variation of style within individual messages as well as such factors as formality level, emotion and specialized or technical language in order to provide translations that reflect the meaning of the source text. While keeping these factors in mind, translations should be literal rather than stylistic in nature. The goal is to produce translations that are true to the source and fluent without veering far from the original text but also without being literal to the point of sounding unnatural.

The following subsections address potential translation issues when dealing with SMS and chat data as source texts. Specific topics include emotion, levels of formality, specialized and technical language, text message language, censored substitution, punctuation, emoticons and URLs. The examples below demonstrate instances in which either a stylistic or a more literal translation might be used. The literal sample translations are the preferred ones for the purpose of this project.

5.11.1 Stylistic Concerns When Translating SMS/Chat Data

Emotions

Translations should express the same degree of emotion as the source text. If a speaker expresses that s/he is infuriated, the target text should use language expressing the same high level of emotion. The translation should express the equivalent of *infuriated* rather than saying "angry."

Furthermore, there may be varying degrees of emotionality, including neutrality, within a single conversation. The target text in this case should express the same range and variety of emotion as the source text. While remaining faithful to the source in regard to choosing appropriate target terms to express emotion, translations should be as close to the original text as possible. For example:

يانهار ابيض، انا ايه الي بيحصلي ده، وقال انا ال خايفة حد يشوفني معاه في الشارع، ويقومو يشوفوني في مدخل العمارة الي فيها الدرس بتاعي كأنه صاحبي وبيوصلني، يادي الفضيحة

```
Literal (preferred)
[Oh my God | what a white day], what's happening to me? I was afraid
somebody might see me with him on the street, but they see me at the
entrance of the building where I take the class, as if he is my
boyfriend and walking. What a scandal!


Stylistic (dispreferred)
[Oh my God | what a white day], what's happening to me? I was afraid
somebody might see me with him on the street, but they spot me at the
entrance of the building where I take the class, as if he is my
boyfriend and walking. Darn it!
```

شايفاكي هتطقي كأنه قالك تعالي نتجوز عرفي، دا واحد كل الي عمله انه اتصل بواحده يعرفها من بعيد بيستهبل، عايز يلم عدد للرحلة الي هو مطلعها.

```
Literal (preferred)
You look like you're going to explode! As if he asked you to marry him
the customary way. All he did was simply calling some distance woman,
```

```
playing fool wanting to get a head count for the picnic he is
organizing.
```

```
Stylistic (dispreferred)
You seem angry! As if he asked you to marry him the customary way. All
he did was simply calling some distance woman, playing fool wanting to
get a head count for the picnic he is organizing.
```

ايه يا مني، انتي خرسا ليه؟ عمرك ماكنتي رخمة كده، فكي شوية

```
Literal (preferred)
Mona, why are you being a mute? You have never been this boorish.
Loosen up!
```

Degrees of formality

Levels of formality may vary. A text conversation may include very formal language, moderately formal language and informal or casual language. In the case of informal or casual language, slang may be present. While degrees of formality should be acknowledged in translation, the translation should not be drastically altered and should follow the meaning of the source as closely as possible.

Specialized and technical language

While most messages will not be technical in nature, some may include technical or specialized language. If technical or specialized terms are used in a message, the target text should include corresponding technical or specialized terms when possible. Highly technical jargon may be difficult for non-specialists, but the translation should not be drastically altered and should follow the meaning of the source as closely as possible.

Borrowings

Dialectal Arabic language contains more borrowings than does MSA. In such cases, translators should preserve the essence of the sentence by translating—or transliterating—these borrowings into English. Examples are provided below:

<ar=1> أعمل **كفر ليتر** متين ، يكون في كل مميزاتك ونقاط قوتك وانجازاتك الدراسية لو يوجد **سي في** ميخرش المية مرتب ومنظم **وبروفيشنل** أوي

```
<en=1> Create a strong cover letter, in which you include all your
qualifications, your strong points and your educational accomplishments
if there is a CV that is flawless, organized and very professional.
```

Semi-Conventional Text Message Spellings

Abbreviations are not as popular in Arabic as they are in other languages, such as English. Dialectal Arabic tends to contain its own abbreviations; moreover, it tends to transliterate certain widely-used English abbreviations and use them to mean the same as in the original language. In such cases, translators should convey the intended meaning of the abbreviations into English. Following are examples:

مس = مع السلامة
```
Translation: good-bye
```

لول = LOL
```
Translation: laugh out loud
```

برب= BRB
```
Translation: be right back
```

تيت = TYT
```
Translation: take your time
```

وب = WB
```
Translation: welcome back
```

باك = back
```
Translation: back
```

Internet/text/chat language

Informal text data may contain a lot of internet/text/chat language, terms, expressions or slang. This poses great difficulty for translation. The rule of thumb for translating such language, terms, expressions or slang is to always translate it into the intended meaning and provide the literal meaning as well when the two are very different, using [text1 | text2] with text 1 being the intended meaning and text2 being the literal meaning.

Repetition

Emphasis in dialectal Arabic is often expressed through repetition of words and expressions. Repetition of such expressions should be preserved in the translation. For example:

<ar=1> **هيجيلك هيجيلك** ورزقك ، وأخيرا أن الارزاق بيد الله بس لازم تكون عارف أولا

```
<en=1> You should know that, regardless, livelihood is in the hands of
God. Your living will come, will come.
```

Furthermore, if a word is repeated in the Arabic source, it should be repeated the same number of times in the English translation. For example:

```
<ar=1> lAlAlAlAlAlA
<en=1> NoNoNoNoNoNo

And not:
<ar=1> lAlAlAlAlAlA
<en=1> NoNoNo
```

NOTE: The above examples concern the repetition of a word. Another frequent occurrence in text and chat messages is the repetition of one or more letters *within a word* for emphasis, for example, as in "Greeeeeeeeat!" When this repetition occurs within a source word, the letters should not be repeated in the English translation. For example:

```
<ar=1> عظيييييم
<en=1> Great!

And not:
<ar=1> عظيييييم
<en=1> Greeeeeeeeat!
```

5.11.2 Emoticons (Emotion Icons)

An emoticon is an ASCII glyph often used to indicate an emotional state in SMS and chat messages. Emoticons should be copied over to the English translation.

The following is an incomplete list of popular emoticons you may see in forum texts:

```
:-)   Standard Smiley (you are joking; satisfied)
:)    Standard Smiley for lazy people
,-)   Winking Smiley. You don't mean it, even if you are joking.
;-)   Winking Smiley. See above.
:->   Follows a really sarcastic remark
```

5.11.3 URLs

URLs that appear in the source text should be copied directly into the translation. For example:

```
<ar=1> عاينوا ...دي عينة لرابط من داخل الموقع
http://www.qassimy.com/vb/showthread.php?t=177072
<en=1> Check this out, a sample link from that website
http://www.qassimy.com/vb/showthread.php?t=177072
```

6   Quality Control at LDC

The quality of a translation is not determined by its style of prose or elegance in use of English but rather mostly in its accuracy. Our definition of "quality" first and foremost requires the translation delivery to be faithful to the source, preserving its original meaning and style. This should be accomplished with the consideration that the translation should also be comprehensible and in fluent English.

Each translation delivery received by LDC is reviewed for completeness, accuracy and overall quality. Payment for completed translation is contingent upon successful completion of the quality review.

Fluent bilinguals working at LDC select a portion of each delivery and grade it according to several criteria. The amount of data selected for review varies depending on the delivery size but at minimum constitutes 1,200 words drawn from multiple documents.

The grading system used by all translation reviewers is outlined below:

Error	Deduction
Syntactic	4 points
Lexical	2 points
Poor English usage	1 point
Significant spelling or punctuation error	½ point (to a maximum of 5 points)
Overlooked file or section	40 points

Below are the error categories:
- o Syntactic: syntactic error
- o Missed translation: lexical error, when a certain word is missing from translation
- o Added translation: lexical error, when a certain word is inserted in translation

- o Wrong translation: lexical error, when translation is wrong
- o Poor English usage: lexical error, when English is awkward
- o Punctuation: punctuation error, when punctuation is changed or missing in translation
- o Spelling error
- o Format problems: spelling/punctuation error, when translation mark-up is missing or incorrect
- o Overlooked file or section: translation missing from a big portion or a file is missing its translation

For each error found, the corresponding number of points will be deducted. For instance, if the original text says *Bush will address the General Assembly of the United Nations tomorrow*, and *tomorrow* is missing in the translation, 2 points would be deducted.

If more than 40 points are deducted from a 1200-word sample, the translation will be considered unacceptable and the whole delivery will be sent back to the translation team for improvement.

If a delivery is sent back to the translation team for further proofreading, the improved version must be completed within 5 business days.

Upon completion of the QC review, the LDC translation team will receive a summary report that includes the following components:

Part 1: Data Profile
Information about the data under review (volume, genre, etc.) and an overall rating for this delivery as *excellent, very good, good, fair* or *poor.*

```
excellent   - 2 or less points deducted
v. good     - 3-10 points
good        - 11-20 points
fair        - 21-40 points
poor        - more than 40 points
```

Part 2: QC Report Summary
Number of words checked:
Error tally: __ points deducted overall
- o Syntactic: __ points deducted
- o Missed translation: __ points deducted
- o Added translation: __ points deducted

- o Wrong translation: __ points deducted
- o Poor English usage: __points deducted
- o Punctuation: __ points deducted
- o Spelling error: __ points deducted
- o Format problems: __ points deducted
- o Overlooked file or section: __ points deducted
- o Other: __ points deducted

Part 3: QC Report Details

For each significant deduction above, at least one example is provided, along with the following information:

- FileID: e.g., `google.com._edfdkfjd_1223.txt`
- Your translation
- Suggested translation
- Discussion: a description of what should be changed and why

7 Guidelines

In case these guidelines prove to be unclear, LDC reserves the right to modify them. Agencies will always use the latest version.