

BOLT Program: Phase 2 Chinese to English SMS/Chat Translation Guidelines

Version 1

July 15, 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.

1	Introduction	3
2	Translation Teams	3
3	File Formats	4
3.1	Source file	4
3.2	Completed translation file.....	4
3.3	File Naming Conventions	5
3.4	Packaging deliveries for LDC	5
4	Delivery of Completed Translations	5
4.1	Email Correspondence about Translation	6
5	Specific Rules for Translation.....	6
5.1	General Principles.....	6
5.2	Proper Names.....	7
5.3	Numbers and Units	8
5.4	Capitalization.....	9
5.5	Punctuation	9
5.5.1	Full-width versus half-width punctuation	9
5.5.2	Missing end-of-sentence punctuation	9
5.6	Idioms	10
5.7	English, Chinese dialect or other language content	11
5.8	Factual Errors in Source Text.....	12
5.9	Typographical errors	12
5.10	Special characteristics of Chinese source texts	13
5.11	Difficult to Translate Source Text.....	13
5.12	Headlines and Titles	Error! Bookmark not defined.
5.13	Poster IDs	Error! Bookmark not defined.
5.14	Special Issues for Translation of Forums.....	14
5.14.1	Stylistic concerns when translating Forum data	15
	Emotions.....	15
	Degrees of formality	15
	Specialized and technical language	16
5.14.2	Internet language	17
5.14.3	Censored substitution.....	18
5.14.4	Emoticons (Emotion Icons)	19
5.14.5	URLs	19
6	Quality Control at LDC	19
7	Guidelines	21
8	Appendix – Commonly Made Mistakes	22

1 Introduction

These guidelines provide an overview of requirements for translation of Chinese text into English to support the BOLT program evaluation of machine translation technology. Genres evaluated in the BOLT program are discussion forum, SMS/chat and telephone conversations. These guidelines will cover SMS and Chat, which will be the genre for Phase 2 of the BOLT program. The data for this phase will be text messages (SMS) and chat (such as Google chat or QQ). Each document will consist of a single SMS or chat conversation between 2 participants (some messages may have been deleted to remove personal identifying information or for other reasons). The length of conversations will vary.

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: a Chinese 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 date/time, message sender, hyperlinks 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 the length of a typical SMS message. This may contain one or more sentences or sentence-like units. Each Chinese segment in the source file consists of 2 components:

<cn=##> is a unique identifier for the Chinese segment;
Chinese 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 Chinese segment.

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

A sample source file appears below:

```
<cn=1>强权在手，什么把戏不能玩转？  
<en=1>  
<cn=2>既然喝的不买，当官饮“路易十三”天价洋酒，也无须自己掏钱。  
<en=2>
```

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 without altering any other part of the file. Altering anything in the Chinese 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 Chinese sentence is translated into multiple English sentences, no blank lines should be inserted between the English sentences.

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

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

<cn=1>强权在手，什么把戏不能玩转？
<en=1> Are there any games they can't play if they are in power?
<cn=2>既然喝的不买，当官饮“路易十三”天价洋酒，也无须自己掏钱。
<en=2> Now that they don't buy any drinks, the officers still don't pay out of their own pocket if they drink super expensive imported spirits like *Louis XIII*.

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_phase2_chinese_set4_NNN.zip [NNN being the name of your agency]

please name the folder containing the translated files
bolt_phase2_chinese_set4_NNN

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

bolt_phase2_chinese_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.Chinese Part 1 - Translation Delivery
AgencyX BOLTPhase2Eval.Chinese Part 2 - Invoice

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

5 Specific Rules for Translation

5.1 General Principles

The goal of the BOLT Translation process is to take Chinese source text drawn from many different genres, both spoken and written, and translate it into fluent English while preserving all of the meaning present in the original Chinese 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 Chinese 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, tone 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.
- 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 Chinese text uses the phrase *Comrade Jiang Zemin*, the translation should **not** be rendered as *Mr. Jiang Zemin*—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 a Chinese proper name has an existing conventional translation into English, that standard translation should be used. For example, 孙中山 is translated into *Sun Yat-sen* instead of *Sun Zhongshan*.

For names without an existing conventional translation, Pinyin should be used in most cases. However, some Taiwanese, Hong Kong and overseas Chinese names do not use Pinyin by tradition, so those names should be rendered in their conventional transliteration. For example, the former Taiwanese president should be translated as *Lee Teng-hui*, not *Li Denghui*.

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

In summary, this is the order of preference in translating Chinese names, within the mainland and outside the mainland:

1. Common transliteration (English name or most commonly used translation of the name).
2. Regional standard for this name, with regional standard determined by context.
3. Pinyin.

When addressing people in Chinese, it may be awkward to directly translate their titles to English. For example, 张老师 can be left as "Teacher Zhang" even if this does not qualify as "appropriate" in English. This cultural consideration applies to all similar titles.

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.

Proper names from languages other than Chinese should be translated into English following the standard conventions for that language. This is particularly important for translating Japanese, Korean and Vietnamese names and also for non-Han Chinese names such as Tibetan, (Inner) Mongolian and Uigur names. Names that originally come from English should be translated using their most common English form. Some

names, although they sound like Chinese names, are not Chinese names at all. For example, 彭定康 (former Hong Kong governor) should be translated as *Chris Patten*, not *Peng Ding kang*.

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 recurring names, 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. Lacking other information to the contrary, translators should simply proceed as if the name were a Chinese 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 and Units

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:

选举名单上有40578个投票人。
There are 40,578 voters on the electoral rolls.

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

大陆有8亿农民，不能把台湾模式硬加在大陆上，这是行不通的。
The mainland has 800 million peasants, so rigidly applying the Taiwanese model to the mainland won't work.

中国现有十三亿人口，发达国家约有十亿人口，让中国全民生活质量达到发达国家水平，其任务超过发达国家总和。
China currently has a population of 1.3 billion. The population of the developed countries is approximately one billion. Improving the quality of life of China's population to the level of developed countries will be a task greater than the task of all the developed

countries combined.

近2万个投票人参加了这次选举。

Nearly *twenty thousand* voters took part in this election.

For units of measurement that may differ between English and Chinese (for example "degrees" or "tons"), the translation should reflect the true number using units of measurement familiar and accepted in English. For example, 吨 should be translated to "metric ton", just as 度 should be translated to "degrees Celsius".

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 geopolitical entities. The first word of each sentence should also be capitalized.

5.5 Punctuation

SMS and chat text, like web text (for example, discussion forums and weblogs), is a very challenging genre in terms of punctuation. Written standards for punctuation vary across languages. As a general rule of thumb, punctuation in the translation should follow standard English punctuation conventions where possible. However, there are specific rules for handling missing end-of-sentence punctuation (see subsection 5.5.2 below). 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.

5.5.1 Full-width versus half-width punctuation

Chinese source text uses standard Chinese punctuation marks, and they need to be converted to their English equivalents. For example, a full-width comma in Chinese should be rendered as a half-width comma in English.

5.5.2 Missing end-of-sentence punctuation

Generally speaking, non-standard punctuation in the source should be preserved in the target (for example, dashes, ellipses, etc.), and translators should **not** standardize punctuation in the translation when it is missing from the source. However, sentences at the end of a segment may be handled differently than sentences elsewhere in the segment. The following describes how translators should approach missing end-of-sentence punctuation at the end of a segment.

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

1. IF the sentence is at the end of the segment (i.e. the final or only sentence of the segment) 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** at the end of the segment (i.e. sentences that precede the final sentence),

THEN:

- Provide standard end-of-sentence punctuation (for example, in the case that a Chinese segment has resulted in multiple English sentences) where possible.

5.6 Idioms

Idioms, colloquial expressions (成语, 习语 refer to <http://baike.baidu.com/view/2990.htm> on definition of 成语) 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.

有传闻，二胖翘辫子了。

There is a rumor that [Kim Jong Il | Fat guy No.2] [**has kicked the bucket** | **'s pony tail has stuck up**].

Intended meaning: has kicked the bucket

Literal meaning: 's pony tail has stuck up

反腐败要想迅速获得效果，要上下齐动员，对腐败分子形成一种“过街老鼠，人人喊打”的氛围，在此基础上建立一种清廉的社会风气，让人民觉得生活在充满正义的天地之中

For anti-corruption efforts to achieve rapid results, there must be an even mobilization at all levels, creating an atmosphere in which corrupt individuals are [**immediately called to attention like a rat crossing the street** | **treated like a rat crossing the street and**

everybody cries “kill it”]. On this foundation, an uncorrupted society can be built that will make people feel that they are living in a heaven-on-earth full of justice.

Intended meaning: immediately called to attention like a rat crossing the street

Literal meaning: treated like a rat crossing the street and everybody cries “kill it”

法院、检察院离政府机关稍微远，有天高皇帝远的感觉，欲所欲为。

The court and the procuratorate are somewhat far from the government organizations, giving a feeling of [**distance from central government control | the heaven being high and the emperor far away**] and action in accordance with one's own will.

Intended meaning: distance from central government control

Literal meaning: the heaven being high and the emperor far away

我没吃过猪肉，还没见过猪跑吗？

[**I’m not a complete neophyte, you know? | I haven’t eaten pork, but haven’t I seen pigs running?**]

Intended meaning: I’m not a complete neophyte, you know?

Literal meaning: I haven’t eaten pork, but haven’t I seen pigs running?

Not all translations of idiomatic expressions need alternatives. If the intended meaning and literal meaning of the idiomatic expression in a context are identical, there is no need to provide alternatives. For example:

黑油窑在武汉之多之久，且高大耸立，排放黑烟遮天蔽日

So many illegal oil kilns have been operating for so long in Wuhan, and they are huge and towering, emitting black smoke **obscuring the sky and the sun**.

武汉黑油窑还要肆无忌惮作孽到何时？

How long will illegal oil kilns in Wuhan continue their **unscrupulous** evil operations?

Be careful not to provide translation alternatives for non-idiomatic expressions. Words/phrases like 老百姓 are commonly used regular expressions, so there is no need to provide alternatives even though the etymology of this word comes from "hundred names."

5.7 English, Chinese dialect or other language content

Occasionally, English, Chinese dialect or another language may appear in the source text. This happens often in forums when internet users post messages in English or other languages or occasionally throw in few words of Chinese dialects, such as Cantonese, Shanghai Hua, etc.

English sentences in the source text should be copied over to the English translation exactly as they appear in the source text. Do not make any changes or corrections to

the English 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, if the source text were the following line:

亚西尔·阿拉法特 (Yasser Arafat) 1929年8月出生于耶路撒冷，是一位逊尼派穆斯林。

It should be translated in the following way:

Yasser Arafat (Yasser Arafat) was born....

Chinese dialect words in the source text should be translated with the intentional meaning. For example:

今天出粮，晚饭我请客。
Today is pay day, so dinner is on me.

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 5.11 for the use of (()). Following is an example:

浦东本地人的私房赔偿拆迁是结棍。。。.
The reimbursement for dismantled personal housing for local Pudong people is ((really big)).

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:

美国总统普京今天访问了莫斯科。
American President Putin visited Moscow today.

汉城将举办2008年奥运会。
Seoul will host the 2008 Olympics.

5.9 Typographical errors

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 add the flag = before the translated word to indicate that it is a correction of a typo. For example:

抗议活动发生在天蓝门广场。

The protest happened in =Tiananmen Square.

连和国需要47亿美元用于人道主义原助。

The =UN needs \$4.7 billion for humanitarian =aid.

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.

5.10 Special characteristics of Chinese source texts

In Chinese informal data such as SMS/Chat, sentences may be grammatically incorrect and elliptical in nature. Translators must be faithful to the original source and stylistically similar, even if this calls for elliptical or ungrammatical English translations. For example:

其中有一副照片上是一个小孩儿，身上穿着很多包装物的那样一个泡沫的东西，看上去非常穷，肯定是无家可归。

One of them is a picture of a young child, wearing a lot of packing materials on his body, something that looks like Styrofoam, looks very poor, definitely has no home to go to.

However, subject omission is an issue occurring in the source which should not be mirrored in the translation. Since subject drop is common in Chinese source texts, yet not grammatically acceptable in English, the subject should be restored in English during translation. For example:

应该说这两国还有相当大的差距。

I should say that there is a relatively large difference between these two countries.

5.11 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:

我们知道，在现在整个岛内，不管是蓝营，还是绿营，还是媒体，还是主流民意，都是一片尊重司法，遵守司法判决的这个结果的一片这样子名义的呼声之下。

We know, that currently, on the whole island, regardless of whether it is the blue camp, the green camp, the media, or mainstream opinion, across the board, they all respect the judiciary, ((and voice the opinion that the result of the judiciary's judgment should be respected)).

我对这种文艺晚会已经厌倦了，特别是国内几年的春节晚会，生硬的拼凑、低俗的搞笑和灌输内容的搀杂

，让我失去了对文艺晚会的兴趣，体内的文艺细胞随着年龄的增长似乎也逐渐的消亡了。

I had become bored of this kind of cultural gala, especially the Spring Festival galas inside the country over the last few years. The awkward way they are slapped together, the lowbrow joke-making and the ((inferiority instilled into the content)) have caused me to lose interest in entertainment galas, while my body's literary and art cells seem to be gradually dying off as my age advances.

Wordplay and puns may also prove difficult to translate. In these cases, a best-guess translation may be necessary as described above. As always, translators should use available resources, including the internet, to find the most appropriate translation for unfamiliar terms or phrases.

5.12 Usernames

For some text message conversations, participants may use nicknames similar to web username or discussion forum poster IDs. For such names, there may be a possible translation into a word or a phrase in English. For example, the name 海洋 may be translated into “Ocean”. It may also be transliterated as “Haiyang”. In cases like this, alternatives should be provided. Specifically, the word or phrase translation should come first, followed by the transliteration:

Format: [translation | transliteration]

NOTE: Translation should always precede transliteration.

Example: 海洋 would be translated in the following way
[Ocean | Haiyang]

5.13 Special Issues for Translation of Informal Language

Translating SMS and chat messages, like discussion forum posts, may prove challenging, as many factors may influence the linguistic style, terminology, grammar and punctuation used in the source text (the message).

The challenge for translators when dealing with SMS and chat messages as source texts is to be aware of the different linguistic styles within such conversations. Furthermore, it is important that translators acknowledge any variation of style 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 informal source texts. Specific topics include emotion, levels of formality, specialized and technical language, internet/text/chat 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.

5.13.1 Stylistic concerns when translating informal language

Emotions

SMS/chat conversations may range from emotionally neutral to emotionally charged—ranging anywhere from slightly emotional to extremely emotional. Translations should express the same degree of emotion as the source text. For example, if a poster expresses that he or she is infuriated, then 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 post. 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:

快跑！要炸了！

Literal (preferred):

Run fast! It will certainly explode!

Stylistic (dispreferred):

Run! It's gonna blow!

宝宝,你要乖乖的听妈妈话,亲亲。

Literal (preferred):

Baby, you must behave and do what I say. Give me a kiss.

Stylistic (dispreferred):

Hon, be good and listen. Kissy-kissy...

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. For example:

天朝CCAV是无原则、无底线、无廉耻的三无宣传机构，它是如此的底俗、庸俗、媚俗。

[CCTV | CCAV] of our country is a three-less propaganda organization which is principle-less, bottom-line-less, shameless. It is so vulgar, philistine and kitsch.

怎么可能？说什么呢？

How come? What did you say?

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.

这里窗户的窗楣并没有显示建筑正里面的装饰风格，而是平面的。

The fenestration here features trabeation displaying none of the ornamentation of the façade, being instead planar.

or

The windows here have lintels that are flat and featureless, with no decoration like those on the front.

Comment: The second rendering makes sense to an average reader, opting for more generalized language, while the original uses specialized terminology. Both translations are acceptable.

说车引擎有450匹,可倍儿难操作。

He says the car's engine has 450 horsepower, but that it is hard to maneuver.

风水地理与传统中医，养生和历法共同拥有两千年哲学和原科学的渊源。

Geomancy shares 2000-year-old philosophical and proto-scientific roots with traditional Chinese medicine, longevity practices, and calendrics.

or

Feng shui shares 2000-year-old roots in early Chinese philosophy and science with Chinese medicine, “nourishing life”, and calendar methods.

Comment:

The second rendering makes sense to an average reader, opting for more generalized language, while the original uses specialized terminology. Both translations are acceptable.

5.13.2 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. Below we list some common Chinese internet language phenomena and examples:

Number substitutions: both intended meaning and literal meaning [intended | literal]

520, 送给所有亲人，兄弟，朋友，想我的，我想的，还有我下一位女朋友！
[I love you | 520]. My love goes to all my family, my brothers, friends, those missing me, those I miss and my next girlfriend!

Comment: 520 = 我爱你

xs56
[LOL | xs56]

Comment: xs56 = 笑死我了

Homophones: both intended meaning and literal meaning [intended | literal]

明明就是**萝卜丝**抓了刘翔的手、什么叫互相的拉拽？还你妹的拳击与动员、这个主持人，你是不是脑子有问题啊？

Obviously it is [**Roberts** | **radish shred**] who grasped Liu Xiang's hand. Where does the push and pull come from? And what is the nonsense of =boxer about? Hey Anchor, are you out of your mind?

Comment: 萝卜丝 literally means radish slice, intended for transliteration of Roberts. There is a typo in “拳击与动员”, so boxer is preceded with = to indicate the typo.

Newly coined words

最近神马**普文二青年**很火啊，但是这个肿么可以代表俺们身在大都市的大好青年们呢？！

These days, [so-called | divine horse] **ordinary, artistic and idiotic youth** are everywhere, but =how can it cover all the youth like us in metropolitan areas?!

Comment: 神马 is homophone of 什么, so both intended and literal translations are provided; 肿么 is an intentional typo of 怎么, so = is used to mark the typo; 普文二青年 is a newly coined word referring to 普通青年, 文艺青年, 二逼青年, translated with only an intended meaning, not with a literal meaning.

Intentional Typos and other errors

看到这个对话 我更羞射了!!肿么办- -

I feel even =shyer after reading this conversation!! =What shall I do - -

Comment: 肿么办 is an intentional typo of 怎么办; 羞射 is an intentional typo of 羞涩; both are marked with = to indicate typo.

Nicknames

A frequent occurrence in forum posts is the use of a nickname to refer to oneself. For example, a poster might call herself 姐, which means *elder sister* although there is actually no familial relationship between her and other posters. In cases like these, please provide translations that are faithful to the Chinese. For example:

咆哮姐

roaring sister *or* roaring elder sister

真相帝

emperor of truth

Unusual transliteration

北鼻 = baby in which 北鼻 literally means north nose

5.13.3 Censored substitution

Occasionally, some SMS/chat data may have certain words/characters redacted by the participants to remove personal information, and such words/characters will appear to be pound signs in the source text. These substitutions need to be copied as they are in the translation.

王##居然也被屏蔽了!

Wang ## was actually censored!

A lot of times, Chinese netizens will come up with various substitutions to bypass censorship. Such substitutions need to be translated into their intended meaning together with the literal meaning with mark-up [intended | literal].

李月月鸟和温影帝比, 谁家更有钱???

[Li Peng | Li Yue Yue Niao] and [Wen Jiabao | Wen the Best Actor award winner], whose family is richer???

看到**CCAV**说回家变容易了后, 再看看**度娘**上说平均刷500次网页才订到一张火车票, 果然是。。。。。(五个字)

After watching [CCTV | CCAV] saying that going home is easier, I read it on [Baidu | Miss Du] that you have to refresh averagely 500 times in order to book a train ticket. No doubt it is (5 words omitted)

5.13.4 Emoticons (Emotion Icons)

An emoticon is an ASCII glyph used to indicate an emotional state in email, news or online posting. 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

5.13.5 URLs

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

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:

- Syntactic: syntactic error
- Missed translation: lexical error, when a certain word is missing from translation
- Added translation: lexical error, when a certain word is inserted in translation
- Wrong translation: lexical error, when translation is wrong
- Poor English usage: lexical error, when English is awkward
- Punctuation: punctuation error, when punctuation is changed or missing in translation
- Spelling error
- Format problems: spelling/punctuation error, when translation mark-up is missing
- Overlooked file or section: translation missing from a big portion or a file

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

- Syntactic: ___ points deducted
- Missed translation: ___ points deducted
- Added translation: ___ points deducted

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

8 Appendix – Commonly Made Mistakes

A1. Overuse of translation alternatives

The use of [|] to indicate translation alternatives should apply specifically to idioms. As mentioned in section 5.6, the format should be as follows:

[intended meaning | literal meaning]

This should not, however, be used merely to indicate translation alternatives—such as synonymous terms or phrases.

For example, the [|] structure is not necessary and should not be used in the following instances:

文革中的一句话：就是要让**老百姓**再受二重罪和苦！！

A saying from the Cultural Revolution: it's letting the [common people | old hundred names] suffer double sufferings!!!

Comment: 老百姓 is not an idiomatic expression. There is no need to provide translation alternatives.

据说中央检查团已经走了，守马路的人也散了，别的不搞好，光守马路有什么用，到处肮脏不堪，地上油污一塌糊涂，**光要面子，不要里子**，

They say that the central inspection team has already gone, the people guarding the roads have also dispersed; if you don't do other things well, what use is just guarding the roads; everywhere is unbearably filthy, the oil contamination on the ground is a complete mess, [they only care about a surface, not the reality | they only want the surface, don't want the inside].

Comment: If translation alternatives are paraphrases, there is no need to provide alternatives.

The following are examples of inferred translations. In these cases, the inferred meaning should not have been included as a translation alternative:

黑暗社会

[unjust society | dark society]

Comment: This is an inference rather than a case of intended meaning and literal meaning. There is no need to provide translation alternatives here.

温水煮青蛙的社会——让百姓安乐死

a society that [starts with warm water to gradually boil frogs | boils frogs with warm water] -- let the people die by euthanasia

*Comment: There is no need to provide translation alternatives here, as the intended meaning is stated in the source (let the people die by euthanasia). **starts with warm water to gradually boil frogs** is an inference of **boils frogs with warm water** and is not an intended meaning.*

我无话可说！！！！湖北这天儿黑啊！

I have nothing to say!!!! [There is absolutely no rule of law here in Hubei! | The sky is dark here in Hubei!]

Comment: The alternatives in the above example are unnecessary. On one hand, 这天儿黑 is not an idiomatic expression. On the other hand, "There is absolutely no rule of law here in Hubei!" is completely inferred.

A2. Missing translation alternatives

In the following example, only a literal translation was provided for the idiom. In this case, both an intended meaning and a literal meaning should have been provided, using the structure [intended meaning | literal meaning]:

人民受多重压迫，比过去的三座大山不知要重多少倍！！！！

The people are suffering many heavy pressures, which are who knows how many times heavier than the **Three Big Mountains** of times past!!!

A3. Inferred translation

The following examples are instances of inference, in which the translation is slightly different than the source meaning. Inference should be avoided.

好好的活着啊，不许闹事！

Stay alive, don't make trouble!

我无话可说！！！！湖北这天儿黑啊！

I have nothing to say!!!! **There is absolutely no rule of law here in Hubei!**

Comment: This is completely inference.

A4. Missing translation and missing/inconsistent punctuation

When translating into English, please be careful not to omit any part of the source. Also, please make sure to retain the punctuation from the original Chinese.

The following is an example of both missing translation and missing punctuation:

《王浩峰聚焦》在一黑油窑现场还吃惊目击到，夏日里，顶着高温艰难地钻进黑油窑内做苦力的是骨瘦如柴年迈老人（67岁），见其已被油污灰土污染成“黑人”。

Wang Haofeng Focus also witnessed surprisingly in a black oil kiln site during summer, a scrawny old man (**67 years**) struggling with the high temperature while squeeze into black oil kilns to do back-breaking work, and he has been turned into a "black man" by the oil waste and dust.

Following are examples of inconsistent punctuation. In both cases, punctuation was added to the English:

都是为了GDP
Everything is just for GDP.

确实在和谐社会只能这样
Indeed in a harmonious society things can only be like this.

A5. Missing markup to indicate typo in the source

In the following example, the word “give” should have been proceeded by a = to indicate a typo in the source:

等青柴大人到啦，会给你升冤的。
When the =imperial envoy comes, he will **give** you justice.

A6. Translation has different meaning than source

In this example, the English does not accurately reflect the Chinese source text:

官员都是老晕花了吧，这么大的烟也没看到。
All the officials have **become dizzy**? They have failed to see even such a big smoke.

A7. Incorrect use of translation alternatives for internet language

The treatment of internet language is discussed in 5.13.2 and is similar to the treatment of idioms in regard to providing translation alternatives. Specifically, the [|] structure is used to indicated intended meaning followed by literal meaning.

In the following example, the alternatives provided are synonymous phrases rather than an intended meaning followed by a literal meaning:

9494, 说不希望有钱了, 但是钱吗, 够花就可以了

Incorrect: [Right, right | it is it is], I said I do not hope to have money, but as for money, as long as there is enough to spend, it will be fine.

Correct: [Right, right | 9494], I said I do not hope to have money, but as for money, as long as there is enough to spend, it will be fine.

DRAFT