

BOLT Program: Arabic to English Translation Guidelines

Version 1

August 7, 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
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	8
5.4	Capitalization.....	8
5.5	Punctuation	9
5.6	Idioms	9
5.7	English or other language content.....	10
5.8	Factual Errors in Source Text.....	11
5.9	Typographical errors	11
5.10	Difficult to Translate Source Text.....	11
5.11	Special Issues for Translation of Speech Sources.....	12
5.11.1	Disfluent Speech	12
5.11.2	Speaker Noise.....	12
5.11.3	Filled Pauses.....	12
5.11.4	Translation of أم, إيه, أه	12
5.11.5	Repetition and Restarts.....	13
5.11.6	Partial Words.....	13
5.11.7	Semi-intelligible and Unintelligible Speech	14
5.11.8	Transcription Mark-ups.....	14
6	Quality Control at LDC	15
7	Guidelines	17
	Appendix: Transcription and translation mark-ups	18

1 Introduction

These guidelines provide an overview of requirements for translation of Arabic text into English to support the BOLT Program evaluation of machine translation technology. These guidelines cover translation of transcribed telephone speech.

This document describes the format of the source text and its translation, and addresses specific issues when translating text from speech genres.

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.

Translators should type the English translation after each “<en=##>” tag without altering any other part of the file. 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.

Speaker IDs in square brackets, e.g. [speaker1], are provided to facilitate clear understanding of conversational speech. They should not be translated or copied over into the translation. Occasionally speaker names do not appear inside square brackets but instead appear at the start of the source text. In such cases, the names should be translated as usual, using the standard guidelines for translating proper names (see Section 5.2).

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=10> [Amr_Moussa] أه أنا علقت وأصدرت بيان رسمي ينعى السيد رفيق الحريري.

<en=10> Uh, I have commented and issued an official statement about the death of Mr. Rafik Hariri.

<ar=11> [Amr_Moussa] أنا شخصيا في غاية الحزن والأسف على فقدان هذا الرجل العظيم اللي كان يعتبر وجها عربيا وليس فقط لبنانيا.

<en=11> I am personally very sad and sorry about the loss of this great man, who was regarded as a prominent Arab personality and not only a Lebanese one.

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.

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:

ABC Agency BOLT Train 01. Arabic Part 1 - Translation Delivery

Agency X BOLT Phase 2 Eval. Arabic Part 2 – Invoice

FGH MT07. Arabic – Schedule Question

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 Translation process is to take Arabic 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 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.

- 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:

أحنا مصنعناش التاريخ لمدة طويلة يمكن توصل لألف وربعميه سنة.

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

انت بتقول 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 اتلشت 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 match the flavor of the punctuation in the source data, while following 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.

Different genres will vary widely in their use of punctuation, and the translation of each genre should respect the flavor of the source text when it comes to punctuation. Transcripts typically have reasonably standard punctuation, which should be preserved in the translation. When transcripts are missing punctuation, it is due to a transcription error, so translators should **add** punctuation in the translation following Standard English punctuation conventions.

Often in transcripts of conversational Arabic, speakers tend to change the subject and restart a sentence in the middle of an unfinished one. When this occurs in the source text, the translation should mimic the source text's punctuation in interpreting the type of pause, restart, or change of subject occurring in speech (see Section 5.12).

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 القطه بسبع أرواح يعني يوقع من البلكونة وما يموتش

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 مزبعة قناة النيل هي زي قمر اربعتاش

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 ضايح. بقالو سنة يدور على شغل ومش لاقى

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 or another language may appear in the source text.

English sentences in 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.

He have a good handwriting.

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.

زار موسكو اليوم الرئيس الأمريكي بوتين.

American President Putin visited Moscow today.

ستستضيف سول الألعاب الأولمبية في عام ألفين وثمانية.

Seoul will host 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,

المعنة التي تمنحها الصين الى افريقيا

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.

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 Speech Sources

This section addresses issues related to translation of transcripts of speech data, such as broadcast news and broadcast conversations (talk shows, call-in shows and the like).

5.11.1 Disfluent Speech

Speakers may stumble over their words, repeat themselves, utter partial words, restart phrases or sentences, and use a lot of hesitation sounds. The sections below describe how to deal with these phenomena in translation.

5.11.2 Speaker Noise

Transcripts may sometimes include markup for speaker-produced noise like coughing, sneezing and laughter. These markers should be copied over into the translation using their original formatting, e.g. {cough}, {laugh}. Please refer to the Appendix for detailed description and examples.

5.11.3 Filled Pauses

Filled pauses are hesitation sounds that speakers employ to indicate uncertainty or to maintain control of a conversation while thinking of what to say next. Filled pauses do not add any new information to the conversation (other than to indicate the speaker's hesitation) and they do not alter the meaning of what is uttered, but they do provide structural information and are an important part of spoken language so they should be translated.

Arabic filled pauses include أه, أم, إيه, أوو etc. They should be translated to their closest counterpart in English, such as *uh*, *um*, *eh* or *ah*.

الولايات المتحدة أه لها دورٌ هام ورئيسي في مسار النزاع العربي الإسرائيلي .

The United States, uh, has an important and principal role in the Arab-Israeli dispute.

أه أم أعتقد إنه، نعم، هناك أمل، وهناك عزم.

Uh, um, I think that, yes, there is hope, and there is determination.

It may sometimes be difficult to decide where to place filled pause in the translated text, since languages vary in where filled pauses occur in speech. The translator is free to shift the location of the filled pause in the translation to make the English more natural, but the filled pause should **not** be deleted or ignored.

5.11.4 Translation of أه, إيه, أم

In conversational speech, أه, أم, إيه can be used in many ways. Translators should differentiate the multiple uses of these words and translate accordingly.

أه, أم, إيه can function as a filled pause. See section 6.12.3 for discussion of how to translate filled pauses.

أه, أم, إيه can also function as a backchannel, in order to provide positive feedback to the speaker to encourage further talk or to confirm that the listener is listening, as in the following example between speaker A and speaker B:

A: أنا أريد في هذه النقطة قبل أن نذهب إلى النقطة الثانية:

B: أه

A: الحكومة الأمريكية لها أهداف إستراتيجية عالمية:

B: أم

A: لها يعني ، عشرات السنين

In backchannel cases, the appropriate English counterpart is something *uh-huh* or *yes*.

5.11.5 Repetition and Restarts

When a speaker repeats him/herself or restarts a sentence halfway through, the repeated words should be translated into English:

هناك مسألتين أساسيتين يحددها، فعلاً، أه ما سوف يؤول إليه الذي تعرض لهذا ال- لهذا الأمر الفظيع.
There are two main issues, defining actually, uh, what will happen to those who were subjected to this %pw, to this horrible thing.

. يعني، ليست ليست إدارة عندما، عندما يطبل عندما يطبل لها بعض الليبراليين المقيمين في أمريكا
I mean, it is not, it is not an administration when some liberals residing in America beat, beat the drum for it.

الليبرالية تأسست في أوروبا.
Liberalism was established in in Europe.

5.11.6 Partial Words

A speaker may stop in the middle of pronouncing a word, which results in a partial word. Transcribers will sometimes use a dash “-” to indicate a partial word in the source text and the point at which word was broken off. Translators should **not** attempt to translate the partial word; their existence should be indicated by using the symbol %pw in the English translation. For instance:

على مستوى المشاعر، جزء كبير بيصاب بق- بقلق شديد.
At the emotional level, a large portion is inflicted with %pw severe worry.

يبدأ يحصل أه أحلام مزعجة جداً، توص-، تصل إلي حد الكوابيس.

They start to have, uh, very disturbing dreams, %pw reaching the point of nightmares.

نتائج ستدل على مستوى عالي من التفاهم الوطني وال- التوافق ال- ال- أه الوطني.

Results that will show a high degree of national mutual understanding and %pw, accord, %pw, %pw, uh, national accord.

هكذا أناك- كمستمع.

This is how I %pw, as a listener.

إذا سمحت، إذا س-

If I may, if %pw

5.11.7 Semi-intelligible and Unintelligible Speech

Sometimes an audio file will contain a section of speech that is impossible to understand. In these cases, transcribers use empty double parenthesis (()) to mark totally unintelligible speech. For example:

قالت الشرطة العراقية إن 37 (()) قد قتلوا.

If it is possible for the transcriber to guess the speaker's words, they transcribe what they think they hear and surround the uncertain transcription/text with double parenthesis. For example:

ضرب زلزال قوي بلغت قوته 8, 6 على مقياس ((ريختر)) نيروبي الساعة 19, 12 بتوقيت غرينتش.

Translators should transfer the double parenthesis over to the English translation, along with the translated words (if there are words to translate). For instance:

قالت الشرطة العراقية إن 37 (()) قد قتلوا.

Iraqi police said that 37 (()) were killed.

ضرب زلزال قوي بلغت قوته 8, 6 على مقياس ((ريختر)) نيروبي الساعة 19, 12 بتوقيت غرينتش.

A strong earthquake, measuring 6.8 on the ((Richter)) scale, hit Nairobi at 12:19 Greenwich Time.

5.11.8 Transcription Mark-ups

Some mark-ups are used in transcripts to mark special speech phenomenon, such as mispronounced words, dialect, foreign language etc.. Some signs need to be copied over into the translation. Please refer to the appendix on how to handle the transcription mark-ups.

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 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 point

Below are the error categories:

- Syntactic: syntactic error
- Missed translation: lexical error, when certain word is missing from translation
- Added translation: lexical error, when certain word is inserted in translation
- Wrong translation: lexical error, when translation is wrong
- Poor English Usage: lexical error, when English is not 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 or incorrect
- 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*, *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.

Appendix: Transcription and translation mark-ups

		external							
symbol	description	status	usage	example	seen by translators?	present in Training/devTest	Present in Eval	translated	in Eval
-	Partial word	required	attached to word at point of truncation	how- or -ver	yes	converted to %pw	yes	yes	yes
--	Incomplete utterance, restart	required	follows interruption or cutoff, surrounded by spaces	blue -- er, green	yes	yes	yes	yes	yes
,	Punctuation, limited to these 4	required	standard usage as in normal writing	, ? , --	yes	yes	yes	yes	yes
(())	Transcriber uncertainty	required	empty	(()) that's what he said	yes	yes	yes	yes	yes
((text))	Semi-intelligible speech	required	surrounds the transcriber's best guess	(()) that's what he said	yes	yes	yes	yes	yes
*	Idiosyncratic words	required	precedes the word	*skalumprious	yes	yes	yes	yes	yes
~	Individual letters	optional	precedes the single letter, used in Chinese only	~A, ~FBI	yes	yes	no	no	no
+	Mispronounced word	required	precedes the word	+Probably (pronounced podably)	yes	yes	no	no	no
{ }	Speaker noise	optional	stand-alone, limited to 4 tags	{cough}, {laugh}, {lipsmack}, {sneeze}	yes	yes	yes	yes	yes
<background>	Non-speaker noise, instantaneous	optional	stand-alone, limited to 1 tag	<background>	no	no	no	no	no
</background>	Non-speaker noise, extended	optional	surrounding word string affected by noise, limited to 1 tag set	<background>text</background>	no	no	no	no	no
<foreign lang="language"></foreign>	Foreign language	required	surrounds foreign text, or empty if transcription unknown	<foreign lang="French"> Bonjour.</foreign>	yes	yes	yes	yes	yes
<non-MSA></non-MSA>	non-MSA speech	required	surrounds non-MSA text, or empty if transcription unknown	<non-MSA> text </non-MSA>	no	no	no	no	no
<foreign lang="non-PTH"></foreign>	non-Putonghua speech	optional	surrounds non-PTH text, or empty if transcription unknown	<foreign lang="non-PTH"> text</foreign>	yes	yes	no	no	no
<telephone></telephone>	telephone speech	optional	surrounds transcription of telephone speech within a broadcast transcript	<telephone> text </telephone>	no	no	no	no	no
[]	indicate that this is a guess based on confusing source text	required	surrounds the translated text	[best guess translation]	added by translator in translation	yes	yes	yes	yes
=	indicate that it is a correction of a typo in the source	required	precedes the translated word	=Africa	added by translator in translation	yes	yes	yes	yes
%pw	partial word in the source	required	precedes the translated word	%pw Africa	added by translator in translation	yes	yes	yes	yes