

**THESE TRANSLATION INSTRUCTIONS ONLY APPLY TO THE
SPECIAL ELICITATION CORPUS OF ENGLISH SENTENCES.**

Introduction

Hello, translators! Thank you for participating in this exercise to learn more about less commonly taught languages (LCTL). From your translations we will be able to learn information about the grammar of your language. This corpus is known as an elicitation corpus; field Linguists have used elicitation for years to learn about new or less studied languages. Traditional elicitation is a collaboration between a speaker of an LCTL and a linguist. The linguist may not speak the LCTL, so a bilingual person like you, known as a language consultant, is needed to communicate with the linguist. The linguist presents sentences to be translated to the language consultant and learns the details of a language by continuously analyzing similar sentences.

Our corpus is inspired by this methodology. In this case you will be translating a series of sentences written in English, but a computer rather than a linguist will present sentences to you. Also, many sentences will contain repeating vocabulary items and each sentence may be only incrementally different from another. This is intended to standardize the elicited sentences and make it easier for linguists or computer programs to compare sentences. Because the computer does not know your language, some of the sentences will seem redundant or not relevant. However, they may be relevant for another language.

As a further note about this guide, we will be using several different languages like Japanese or Spanish, as examples to show the right or wrong way to translate using the elicitation corpus. They are almost certain to be different from the language you will be translating and none will be LCTLs. They are intended to be illustrations only.

Translating Sentences

Through the process of elicitation we will present you with a series of sentences and context fields. Each sentence should be considered as an independent utterance. That is, we should be able to give you the elicitation sentences in any order and the translation should be the same. In figure 1 we have two unrelated sentences. In a regular conversation if we heard those two sentences in succession we would assume that the person Simon forgave was Frank. However, in the elicitation corpus both of these sentences should be considered independently. Your translation of 1a and 1b should be the same no matter in what order the sentences appear and the information in one sentence does not apply to any other. It would be best to think of the Simon in 1a as a different, unrelated person to the Simon in 1b in a different, unrelated situation.

1 a.

Sentence: Frank hit Simon.
Context: n/a

1 b.

Sentence: Simon forgave him.
Context: n/a

Figure 1: Each sentence is an independent utterance. The sentences are **not** a related sequence of events. These two sentences should be translated as separate, unrelated events. That is, the information in 1a should not influence the translation of 1b and vice-versa.

Additionally, we are looking for the most natural, consistent translations possible. We want to study your language as it is used naturally. However, here may be some vocabulary words or utterances that do not translate exactly into your language. Please use the most natural equivalent possible and use it consistently. Consistent translations will help us draw accurate conclusions about the grammar the LCTL in question. Looking at the examples in figure 2 we can see two errors. First, while the Japanese word 'Jidouhanbaiki' literally means 'Automatic sales machine' (as seen in 2 a), the phrase 'vending machine' used in 2b is much more natural than. For our purposes it is much better to use the natural translation over the literal one when there is an equivalent, natural phrase existing in the translation language. 'Vending Machine' carries the same amount of information to an English speaker as 'automatic vending machine', but manages to be clearer.

Secondly, now that we know that the better translation is 'Vending Machine', this phrase should be used whenever we see 'Jidouhanbaiki'. Uniform translation is important in order to compare sentences and learn about your language's grammar.

2 a

Sentence: Jidouhanbaiki ga kowareta.
Context: n/a
Translation: The automatic sales machine broke.

2 b

Sentence: Jidouhanbaiki go kowareru.
Context: n/a
Translation: The vending machine will break.

Figure 2: Words and phrases should be translated with the most natural phrase available and translated consistently. The above example with translations from Japanese to English uses inconsistent translation in 2a and 2b and chooses an awkward turn of phrase in 2a.

There may be multiple natural sounding translations for any given utterance in your language. We can only receive one translation for each sentence, so you must choose the one that you would be more likely to use.

3a.

Sentence: She is a teacher. Context: n/a Translation: Ella es una Maestra.
--

3b.

Sentence: She is a teacher. Context: n/a Translation: Es una Maestra.

Figure 3: Two translations may be equally correct for a given sentence. The above English sentence can be translated several ways in Spanish, which is a pro-drop language. That is, the sentence can be translated with or without a subject. The decision of which sentence will be used is left to the translator. However, if they decide to leave off the subject, they should do the same with all similar sentences if they can be translated that way. The same is true in the case where the translator decides to include the subject. Consistency is important.

Also, as mentioned above, our English sentences will use repeating vocabulary and structure. Over the course of the corpus, some of the sentences will seem redundant or not relevant. However, they may be relevant for another language. It is perfectly acceptable to translate some sentences identically for your language. Again, consistency is encouraged. In figure 4 we see that the same English sentences translate differently in to Spanish (4a and 4b), but the same in Tagalog (4c and 4d). If you will be translating into a language that does not mark gender like Tagalog you would translate those sentences the same. It is possible to explicitly state that the teacher is 4c is male, but for our purposes this is not ideal. To do so would change the grammatical structure of the Tagalog translation. Again, be consistent with the way you translate each utterance.

4 a.

Sentence: She is a teacher. Context: n/a Translation: Ella es una Maestra.
--

4 b.

Sentence: He is a teacher. Context: n/a Translation: Él es un Maestro.
--

4 c.

Sentence: She is a teacher. Context: n/a Translation: Guro siya

4 d.

Sentence: He is a teacher. Context: n/a Translation: Guro siya
--

Figure 4: A comparison between two sets of sentences translated into Tagalog and Spanish. In Spanish translations 41a and 4b the gender difference on the subject is evident in two three places: the subject, the predicate nominal and on the dependent of the predicate nominal. In the Tagalog translations 2a and 2b there is no difference at all.

How to use the context field

The context field is intended only to provide information that is not encoded in the English sentence. However, You will be translating the English sentence only. Because this corpus was not designed with any one language in mind the context field may or may not influence how you translate a particular utterance. For example, in figure 5 depending on how number is marked in a particular LCTL the two translated sentences may be identical. The Arabic translation (5a and 5b) incorporates the information in the context field, while the Spanish translation (5c and 5d have no need for it).

5 a.

Sentence: You wrote. Context: You = five men Translation: antum katabtum
--

5 b.

Sentence: You wrote. Context: You = two men Translation: antumaa katabtumaa

5c

Sentence: You wrote. Context: You = five men Translation: escribieron

5d

Sentence: You wrote. Context: You = two men Translation: escribieron
--

Figure 5: The two sentences translated into Modern Standard Arabic (5a and b) are translated differently based on the number of people 'You' represents. However, the Spanish translations remain the same in 5c and 5d.

With respect to context fields you should consider how the additional information influences your translation. Don't feel obligated to include the information given in the context field in the sentence translation. For example, in figure 6 the information source of a sentence may influence translation in some languages. There are languages will translate an utterance differently based on whether the speaker witnessed the event first hand or heard it from someone else. It is important to consider how this difference is manifested in your language. Would you translate 'Bill ate the cake' differently if the speaker saw it versus heard about it? If the answer is yes, then you should incorporate the information in the context field into your translation. If no, then both sentences should be translated identically.

- 6 a.

Sentence: Bill ate the cake.
Context: The speaker saw Bill eat the cake
- 6 b.

Sentence: Bill ate the cake.
Context: The speaker heard about Bill

Figure 6: Two sentences that differ by information source

Speaker vs. Listener

There will be references to the 'Speaker' and 'Listener' in this document. The speaker is the originator of a given utterance and can be considered to be 'first person'. If a gender is not indicated the speaker may be taken to be an adult male. However, please be aware that the speaker will change gender throughout the corpus and these changes will be indicated in the context field.

Likewise, the listener is the entity to which the utterance is spoken. This is equivalent with second person. The listener should also be interpreted as an adult male unless otherwise specified.

The speaker and listener should also be considered to have equal rank and age unless otherwise noted. For languages where solidarity is important, both should be considered to have the familiarity of co-workers.