

An Assessment of Language Elicitation without the Supervision of a Linguist

Alison Alvarez, Lori Levin, Robert Frederking
[nosila | lsl | ref]@cs.cmu.edu
Jill Lehman jill@kidaccess.com

Language Technologies Institute
Carnegie Mellon University

Abstract

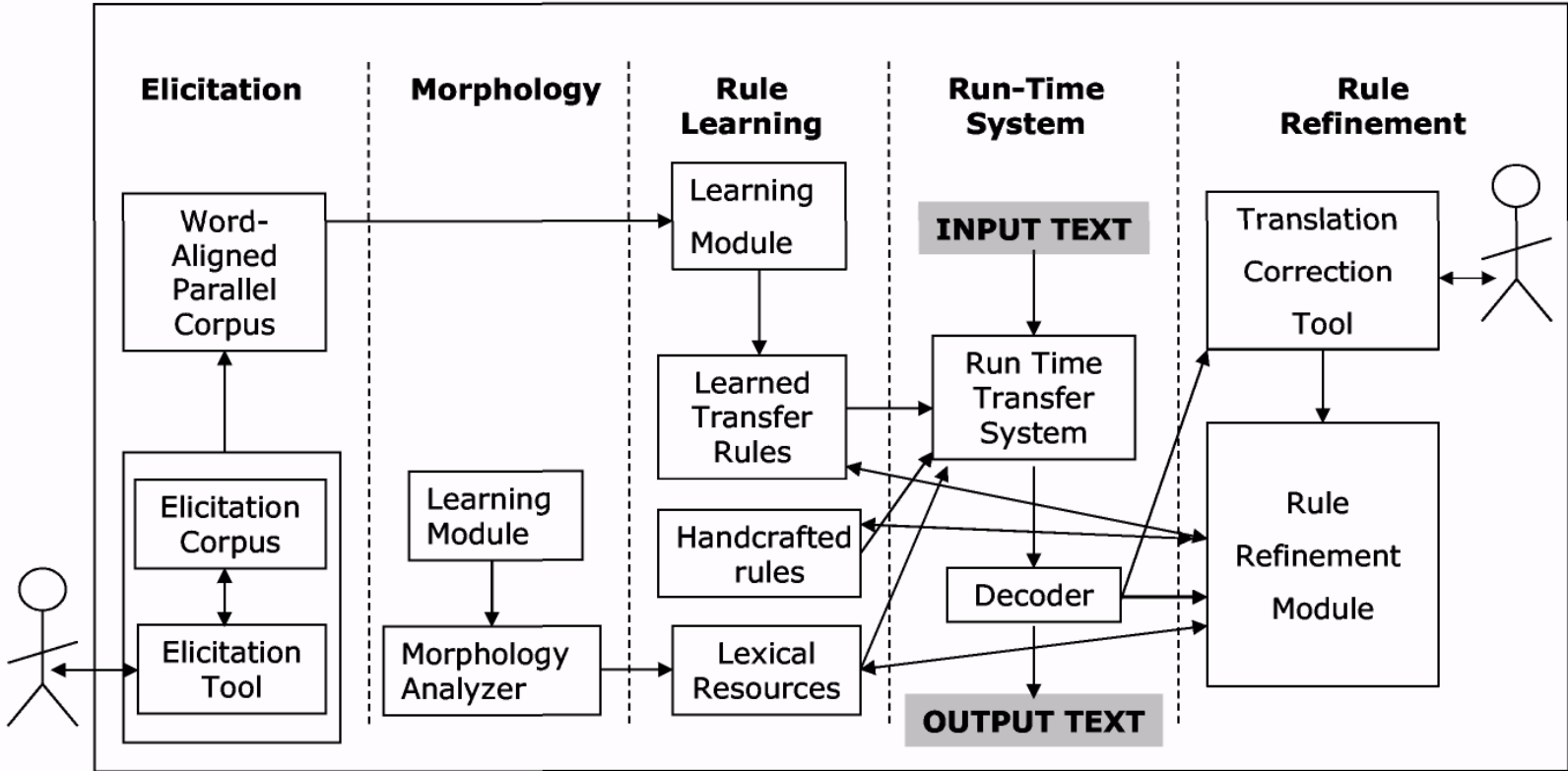
The **AVENUE** machine translation system is designed for *resource poor scenarios* in which parallel corpora are not available.

In this situation, parallel corpora are created by bilingual consultants who translate an *elicitation corpus* into their languages.

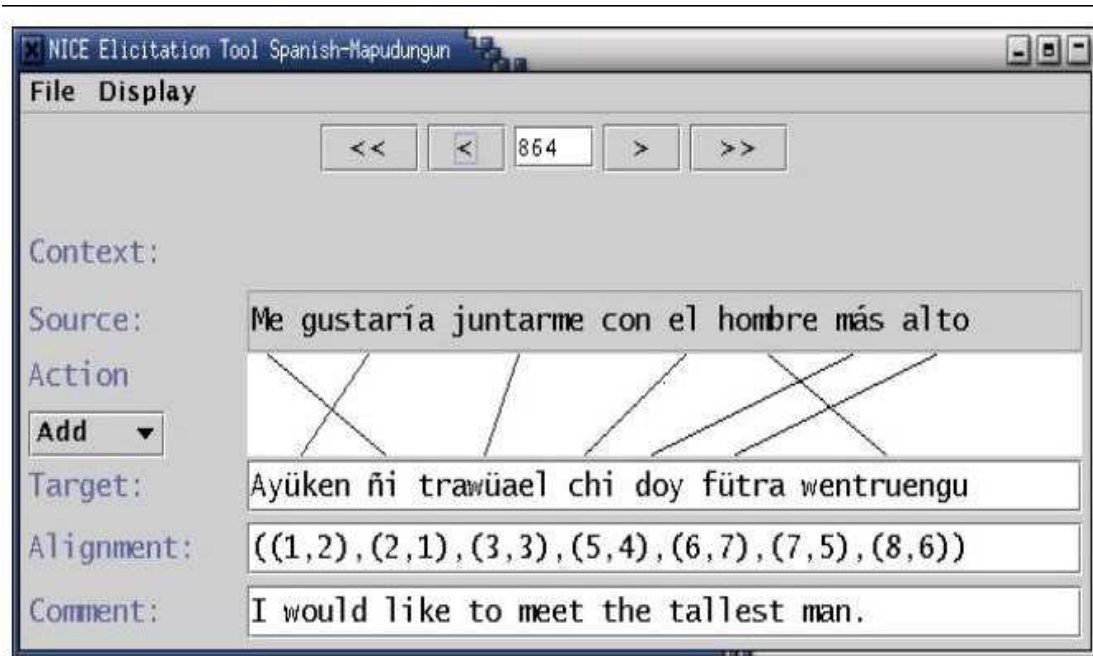
This paper is concerned with *evaluation* of the elicitation corpus: is it suitably designed so that a bilingual consultant can produce reliable data without the supervision of a linguist?

We evaluated two translations of the elicitation corpus, one into *Thai* and one into *Bengali*. Two types of evaluation were conducted: an *error analysis* of the translations produced by the Thai and Bengali consultants, and a comparison of *Example Based MT* trained on the original human translations and on corrected translations.

Avenue Architecture



AVENUE Elicitation Tool



(Language pair shown is Spanish/Mapudungun.)

Linguistic Resource: REFLEX

As part of a U.S. government project called REFLEX, we produced an elicitation corpus of 3124 English sentences, which the Linguistic Data Consortium (LDC) is translating into a number of languages, beginning with Thai and Bengali. Contrary to the AVENUE scenario, no hand alignments were done, and there was no supervision of the translators by the AVENUE team.

Elicitation Corpus: example sentences

- Mary is writing a book for John.
- Who let him eat the sandwich?
- Who had the machine crush the car?
- They did not make the policeman run.
- Our brothers did not destroy files.
- He said that there is not a manual.
- The teacher who wrote a textbook left.
- The policeman chased the man who was a thief.
- Mary began to work.

Elicitation Corpus: detailed example

srcsent: We baked cookies.

context: We = 5 men;

```
((actor ((np-function fn-actor) (np-general-type pronoun-type)(np-person person-first)
  (np-identifiability identifiable) (np-pronoun-exclusivity inclusivity-neutral)
  np-number num-pl) (np-biological-gender bio-gender-male)(np-animacy anim-human)
  (np-specificity specific)(np-pronoun-antecedent antecedent-not-specified) (np-distance
  distance-neutral)))
(undergoer ((np-function fn-undergoer)(np-person person-third)(np-identifiability unidentifiable)
  (np-number num-pl)(np-specificity non-specific)(np-animacy anim-inanimate)
  (np-biological-gender bio-gender-n/a)(np-general-type common-noun-type)
  (np-pronoun-exclusivity inclusivity-n/a)(np-pronoun-antecedent antecedent-n/a)
  (np-distance distance-neutral)))
(c-polarity polarity-positive) (c-v-absolute-tense past) (c-v-lexical-aspect activity-
accomplishment)(c-general-type declarative-clause)(c-my-causer-intentionality intentionality-
n/a)(c-comparison-type comparison-n/a)...
```

Figure 1: An abridged feature structure, a source language sentence and its context field

Minimal Pairs: Change vs. No Change

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

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

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

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

Figure 2: Context information isn't always incorporated into target language translations. The two sentences translated into Modern Standard Arabic (2a and 2b) are translated differently based on the number of people 'You' represents. However, the Spanish translations remain the same in 2c and 2d. This example and further ones can be found in our translator guide.

Elicitation Error Analysis: statistics

Thai Elicitation Errors		
Source Sentence Over-Translation	845	79.41%
Context Over-Translation	57	5.35%
Under-translation	88	8.48%
Mistranslation	68	6.39%
Grammar Mistakes	6	0.19%
Total	1064	100%

Bengali Elicitation Errors		
Source Sentence Over-Translation	0	0.0%
Context Over-Translation	24	6.68%
Under-translation	5	1.39%
Mistranslation	76	21.17%
Grammar and Spelling Mistakes	254	70.75%
Total	359	100%

Figure 3: Total elicitation errors for the Thai and Bengali translations of the elicitation corpus.

Elicitation Error Analysis: detailed examples

a. Context Over-translation								
Bengali target:	বিজয়া কয়েক সপ্তাহ আগে বংকিমকে বইগুলি দিচ্ছিল.							
transliteration:	BAiJAYYAaa	KAYYAeKA	SAPAVIRTAaaHA	AAG Ae				
	BAANUKAiiMAKAe	BAIGAuLai	DAiCAVIRCHAiLA.					
gloss:	Bijoya	a-few	moment-plural	before				
	Bankim-acc	books-plural	give/third-person/progressive					
source:	Bijoya was giving Bankim books.							
context:	Translate this sentence as if the incident it refers to happened minutes ago.							
b. Source Sentence Over-translation								
Thai target:	ผู้ชาย คน นั้น มี ความสุข							
transliteration:	pôo chaai	kon	nán	mee	kwaam sòok			
gloss:	man	person	that	is	happy			
srcsent:	The man was happy.							
context:								
c. Under-translation								
Thai target:	ผู้ชาย คน นั้น จะ ตำหนิ เด็กผู้หญิง คน นั้น							
Transliteration:	pôo chaai	kon	nán	jà	dtam-ni dèk	pôo ying	kon	nán
gloss:	man	person	that	will	reprimand	girl	person	that
srcsent:	The man will criticize the girl.							
context:	Translate this as if the speaker heard this information from a rumor .							
d. Mistranslation								
Thai target:	รั้ว รอบ ทุ่งหญ้า พังทลาย ลง							
Transliteration:	rúa	rôp	tông yâa	pang tá-laai	long			
gloss:	fence	around	pasture	fall	down			
srcsent:	The fence around the pasture collapsed.							
context:								
e. Spelling and Grammar Mistakes								
Bengali target:	মহিলাটি যে গুদামে নয় কথা বলিতেছে.							
Transliteration:	MAHiLaaTTi	Ye	GAuDAAA MAe	NAYYA	KATHAaa			
	BALAiTAeCHAE.							
gloss:	woman-def	what	store	negative	statement			
	talk/third-person/progressive							
srcsent:	The woman who is not in the store is talking.							
context:								

Figure 4: This figure catalogs examples of our five types of elicitation errors.

EBMT Thai/English Experiment

Compare EBMT trained on original REFLEX data against EBMT trained on corrected sentences; see what effect corrections have on BLEU score of resulting EBMT system. (EBMT being used as a stand-in for the eventual learned transfer-based system.)

- 2924 training sentence pairs
- 100 tuning sentence pairs
- 100 test sentences (always from corrected set)
- Same split in both data sets
- English Language Model trained on other data

EBMT BLEU Results	
Uncorrected Thai	0.499
Corrected Thai	0.552

This is a **9.6%** relative improvement.

Discussion

The BLEU scores reported here are higher than normal for several reasons, primarily the shortness and redundancy of the sentences in our corpus. Since we are primarily interested in the difference between the datasets, this is not a major problem.

Conclusions

From error analysis:

Improvements are possible in the process:

- Current documentation could be clearer, and use more examples. Could explicitly teach about tension between natural and faithful translations.
- Corpus sentences could be less unwieldy, be provided in a discourse context, and include visual aids.
- Training should be provided, with a pre-test and detailed feedback.

From EBMT experiment:

Elicitation errors *significantly* affected the performance of the EBMT system. However, despite this, the Bleu score declined by only **9.6%**, providing some evidence that the *uncorrected* translations would still be able to train a usable system.

Acknowledgements

We would like to thank our language consultants, Dipanjan Das, Satanjeev Bannerjee and Vorachai Tejapaibul. In addition, we would like to thank Aaron Phillips for his help building and testing our Example Based Machine Translation system.

Elicitation Error Analysis: discussion

Thai over-translation: Thai does not mark definiteness. Thai translator improperly used “that” 578 times (out of 845 over-translations) to try to mark definiteness. Fixing this reduces total elicitation error for Thai by 68%.

Bengali non-native errors: We believe the Bengali translator was *not a native speaker*. Example 4e should be “The woman who is angry, she is talking”. Inanimate markers were used on animate noun phrases. The popular name “Bankim” was misspelled. These sorts of errors accounted for 845 (71%) of the Bengali errors, versus only 6 such errors in the Thai data.