

11-731: Machine Translation

Homework Assignment #1:

Language Divergences and their Impact on Machine Translation

Out: Wednesday, January 14, 2009

Due: Wednesday, January 21, 2009

For the purpose of this assignment, we assume you have some working knowledge of at least one language other than English. You can work with any non-English language of your choosing, or use examples from several different languages if you wish to do so.

In class, we surveyed several different classes of language divergences and examined their impact on machine translation. The main divergence classes that we examined were the following:

- Lexical Differences
- Structural Differences
- Differences in Mappings of Syntax-to-Semantics
- Constructions and Idioms

For each of the above types of divergences, find and locate **at least one** (but possibly more) concrete example of this divergence type in **the foreign (non-English) language** of your choice, which was **not already shown in class**. For each such example:

1. Show the correct translation of your example **into English**, and also give a literal gloss translation of your example.
2. Briefly explain (in a few short sentences) why you believe this example is of the specific divergence type that you indicate it demonstrates.
3. Create a full sentence in the foreign language that contains your example, and translate this sentence into English using “Google Translate”. Show the resulting translation.
4. Briefly analyze the resulting Google translation outcome. Did Google translate this sentence correctly? If not, why not, and is this a direct consequence of the divergence type that you indicated? If the Google translation is correct, why do you believe Google was able to correctly overcome this divergence?

We recommend that you locate divergence examples by examining actual real text in the foreign language over the web. If you do so, please explicitly identify the source of the text. If you come up with the example on your own, search for it on the web and give some indication of whether this example is frequent or prevalent in your language.

Please submit the write-up of your homework electronically by email to the class TA, Sanjika Hewavitharana <sanjika@cs.cmu.edu>, by class time on Wednesday, January 21. Questions should be addressed by email to Alon Lavie <alavie@cs.cmu.edu>.