



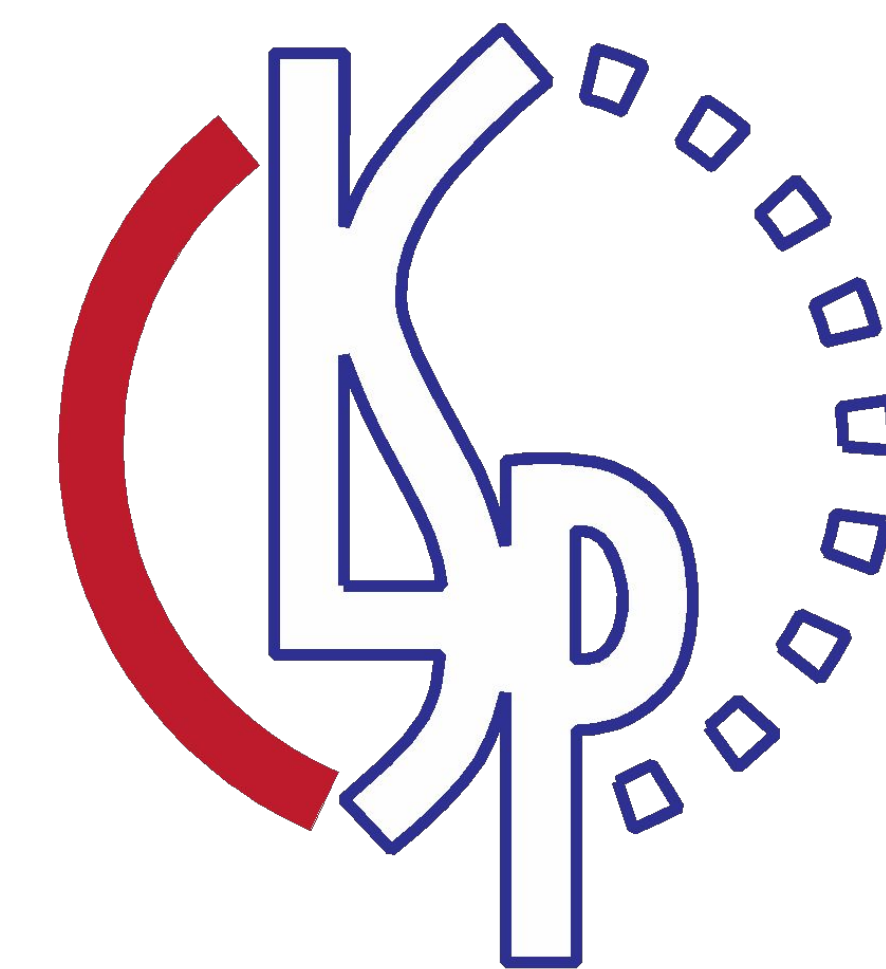
An Interactive Teaching Tool for Introducing Novices to Machine Translation

Huda Khayrallah, Rebecca Knowles, Kevin Duh, Matt Post

Department of Computer Science

Johns Hopkins University

{huda, rknowles}@jhu.edu, {kevinduh, post}@cs.jhu.edu



Abstract

The first step in the research process is to develop an understanding of the problem. Novices may be interested in learning about machine translation (MT), but often lack experience and intuition about translation (either by human or machine) and its challenges. This tool allows students to interactively discover why MT is an open problem by having them recapitulate some of the ideas explored by the first machine translation researchers in the 1940s, and encourages them to ask questions, propose solutions, and test intuitions. In this hands-on activity students build and evaluate their own MT systems using curated parallel texts.

Lesson Overview & Instructions

Setup: students receive some parallel text for English and a recently-discovered “alien language” (actually ciphered Spanish). Their task is to produce a machine translation system to translate from the new language into English.

Challenge: their system will be judged not only on its ability to correctly translate the known parallel text, but also on its ability to translate *novel texts*.

Instructions: student groups are given the following instructions for the tool and can work together or compete against one another to build the best translation systems.

Instructions

Adding rules

To enter a rule, type its source and target sides in the boxes above. For example, you might enter:

source="ne peux pas" target="can't"

Then click 'Add Rule' to have the rule added. The rule will appear in the list above, and any data in the input box on the first tab will be retranslated.

For more advanced usage, you can add hierarchical rules. These are rules with a variable inside them. The variable serves as a substitution point for the source and target sides. For example:

source="ne X pas" target="do not X"

X here will be replaced with another phrase pair that you have entered. If you have another entry for "veux" and "want", for example, the decoder can use these two rules to translate the source phrase "do not want".

Up to two variables are permitted: X and Y.

Building Systems & Intuition

Parallel Text Data	
Source	English
Jesa Q . vod oxahod !	Look at the bees !
O Joseeo va fwdro jesos vod oxahod .	Mary likes to watch the bees.
Va fwdro qasvod sagifas vo jeav .	She likes to see them gather honey.
Bi vo negob .	They do not sting her.
O avvo va fwdro ouwposvod .	She likes to help them.
...	...

Students examine the parallel text to determine translations of words and phrases. They then add new Hiero-style rules (Chiang, 2005) to the system. These can be:

Rule Type	Example
word-to-word	Joseeo → Mary
phrase-to-phrase	vod oxahod → the bees
hierarchical (X)	X bafsi → black X
hierarchical (X & Y)	O X va fwdro Y → X likes Y

Students can also remove rules (by clicking the red X next to each rule).

Rules

X bafsi black X Add Rule

Source	Target
X oxahod	bees
X givjabod	hives
X u	and
X vo	the
X vod	the

Input Rules Parameters Instructions

Rules

X va fwdro Y Y likes X Add Rule

Source	Target
X givjabod	hives
X oxahod	bees

Output

Jesa Q. vod bees!

Jesa Q. giji qwavob o dwd hives.

Sagifab vo jeav pa vod mvisad u vo nibab ab dwd hives.

O Joseeo va fwdro jesos vod bees.

Va fwdro qasvod sagifas vo jeav.

As they build rules, students begin to encounter a number of translation challenges, including but not limited to: word ordering, alignment, lexical ambiguity and word sense disambiguation, morphology, and more.

Usage & Learning Outcomes

Students can examine the effects of adding rules by using their system to translate the known parallel text.

Input Rules Parameters Instructions

Input

Load text from: Text 1

Jesa Q . vod oxahod !
Jesa Q . giji qwavob o dwd givjabod .
Sagifab vo jeav pa vod mvisad u vo nibab ab dwd givjabod .
O Joseeo va fwdro jesos vod oxahod .
Va fwdro qasvod sagifas vo jeav .
Bi vo negob .

Translate

Output

Look the bees!

Look giji qwavob o dwd hives.

Sagifab the jeav pa the mvisad u the nibab ab dwd hives.

By hand-engineering MT system rules, which they can evaluate on real data, students gain intuition about why early MT research took this approach, where it fails, and what features of language make MT a challenging problem even today. This interactive exercise can be augmented by a discussion of state-of-the-art MT techniques and challenges, focusing on areas or aspects of linguistic complexity that the students found difficult.

We have used this as part of a two week summer school on natural language processing. The morning was spent on the interactive lesson in teams, periodically updating students on the performance of their systems (BLEU on training and held-out data); the afternoon was spent discussing and learning how the challenging parts of the activity relate to open questions in the field. The lesson could be used as a standalone activity or as part of a larger AI or NLP course.

Learn More!

For more info and a video demo, visit:
<http://www.cs.jhu.edu/~huda/MTTeachingTool>