CSE 5525 Homework 4: Machine Translation
(Optional Extra Credit)

Alan Ritter

In this assignment you will train phrase-based translation models using Moses\(^1\), a popular open-source phrase-based machine translation system and GIZA++\(^2\) a widely-used implementation of the IBM word alignment models. First you will build a French \(\rightarrow\) English translation system using a (relatively small) parallel corpus of 130,000 sentence pairs. Next you will build a system that “translates” from standard English into Shakespearean style using provided parallel text that has been scraped off the web.\(^2\)

1 Installing Moses

First, install Moses by following the instructions here (using Linux is recommended): http://www.statmt.org/moses/?n=Development.GetStarted

2 Installing GIZA++

Next install GIZA++:

```
    git clone https://github.com/moses-smt/giza-pp.git
    cd giza-pp
    make
```

Now, copy the GIZA++ binaries to where Moses can find them, for example:

\(^1\)http://www.statmt.org/moses/
\(^2\)http://nfs.sparknotes.com/romeojuliet/page_4.html
cd ~/mosesdecoder
mkdir tools


3 Building a Phrase-Based French → English Translation System (2 points)


Pick a few French sentences to translate into English and include the output in your writeup.

4 Building an English → Shakespeare System (3 points)

Now build a system to paraphrase normal English into Shakespearean style using the provided parallel text in the resources page on Piazza. Use Shakespeare’s plays to build a language model and the provided parallel text for the translation model. Try running some lines from modern movie scripts$^3$ through your translation system and find some fun examples to include in your report. (Note: you may want to tokenize the input sentences before inputting them into your system. It’s possible to do this by hand, but if you are translating a lot of text you might want to use the provided tokenization scripts)

4.1 Word Alignments (1 point)

Inspect the file containing IBM word alignments, for example, like so:

```
> zless original-modern.A3.final.gz
```

Find at least one sentence pair to include in your report where the alignment is mostly correct, and one pair where there are obvious errors.

---

4.2 Phrase Table (1 point)

Have a look at what’s in the phrase table, for example using the following command:

> zless phrase-table.gz

Again, include a few examples of phrase-pairs that make intuitive sense, and some that appear to be incorrect.

What to Turn In

Please turn in the following to the dropbox on Carmen:

1. A brief writeup that includes the examples requested above (there is no need to turn in any code for this assignment).