

LING 120:
Language and Computers
Semester: Fall '17

Instructor: Sowmya Vajjala

Iowa State University, USA

1 December 2017

Today's class

- ▶ Reminder about Finals and the related deadlines
- ▶ Discussion about questions from last class
- ▶ Evaluation of MT
- ▶ Conclusion of MT

Discussion about question from last class

- ▶ Task: identify English-Indonesian word translations based on the provided comparable article between the two languages.
- ▶ Solution: `http://nacloweb.org/resources/problems/2011/FS.pdf`
- ▶ Strategy: Not looking for exact meanings (we don't know the other language!). Looking for contexts (surrounding words), frequency of occurrence, words that go together etc.

Discussion about question from last class

Different ways of machine translation:

- ▶ Example based MT: Look for matching words or phrases in previous human translations, and assist human translators by doing this.
- ▶ Direct MT: Word by word generation of translation
- ▶ Syntactic Transformation based MT: Rules for changing syntactic structure between source and target languages
- ▶ Combining Direct and Syntactic MT
- ▶ Statistical MT (learning from data)

Questions before I start with evaluation

- ▶ What is a good way to evaluate MT?

Questions before I start with evaluation

- ▶ What is a good way to evaluate MT?
- ▶ Is doing human evaluations easy or difficult?
- ▶ If I don't have humans sitting there to do evaluations, how can I do automatic evaluations?

Questions before I start with evaluation

- ▶ What is a good way to evaluate MT?
- ▶ Is doing human evaluations easy or difficult?
- ▶ If I don't have humans sitting there to do evaluations, how can I do automatic evaluations?
- ▶ What aspects should be considered while doing evaluation of MT systems?

Questions before I start with evaluation

- ▶ What is a good way to evaluate MT?
- ▶ Is doing human evaluations easy or difficult?
- ▶ If I don't have humans sitting there to do evaluations, how can I do automatic evaluations?
- ▶ What aspects should be considered while doing evaluation of MT systems?
- ▶ Why bother about evaluation, really?

Questions before I start with evaluation

- ▶ What is a good way to evaluate MT?
- ▶ Is doing human evaluations easy or difficult?
- ▶ If I don't have humans sitting there to do evaluations, how can I do automatic evaluations?
- ▶ What aspects should be considered while doing evaluation of MT systems?
- ▶ Why bother about evaluation, really?
- ▶ How are human translations evaluated in professional translators' world?

Questions before I start with evaluation

- ▶ What is a good way to evaluate MT?
- ▶ Is doing human evaluations easy or difficult?
- ▶ If I don't have humans sitting there to do evaluations, how can I do automatic evaluations?
- ▶ What aspects should be considered while doing evaluation of MT systems?
- ▶ Why bother about evaluation, really?
- ▶ How are human translations evaluated in professional translators' world?

Going backwards with answers-1

- ▶ Q: How are human translations evaluated in professional translators' world?
- ▶ A: Quality is evaluated in terms of number of words edited/corrected in post-translation editing and proof reading.
- ▶ Q: Can we apply the same measure to MT?

Going backwards with answers-1

- ▶ Q: How are human translations evaluated in professional translators' world?
- ▶ A: Quality is evaluated in terms of number of words edited/corrected in post-translation editing and proof reading.
- ▶ Q: Can we apply the same measure to MT?
- ▶ A: May be not sufficient. First, MT may always need editing, second: writing patterns of a human and a machine can be completely different. We may need more evaluation measures along with this.

Going backwards with answers-2

- ▶ Q: Why bother about evaluation, really?
- ▶ A: Some example scenarios:
 - ▶ For end users: should I really use MT on my website? Is it any good, or will it reduce my customer base because it will frustrate them?

Going backwards with answers-2

- ▶ Q: Why bother about evaluation, really?
- ▶ A: Some example scenarios:
 - ▶ For end users: should I really use MT on my website? Is it any good, or will it reduce my customer base because it will frustrate them?
 - ▶ For MT software developers: How do I know how well are new changes in my program improving the translation quality?

Going backwards with answers-2

- ▶ Q: Why bother about evaluation, really?
- ▶ A: Some example scenarios:
 - ▶ For end users: should I really use MT on my website? Is it any good, or will it reduce my customer base because it will frustrate them?
 - ▶ For MT software developers: How do I know how well are new changes in my program improving the translation quality?
 - ▶ For researchers: How do we compare different MT systems around the world, made for different reasons, by different people, following different approaches?

Going backwards with answers-2

- ▶ Q: Why bother about evaluation, really?
- ▶ A: Some example scenarios:
 - ▶ For end users: should I really use MT on my website? Is it any good, or will it reduce my customer base because it will frustrate them?
 - ▶ For MT software developers: How do I know how well are new changes in my program improving the translation quality?
 - ▶ For researchers: How do we compare different MT systems around the world, made for different reasons, by different people, following different approaches?

Going backwards with answers-3

- ▶ Q: What aspects should be considered while doing evaluation of MT systems?
 - ▶ A: many
 - ▶ Quality (naturalness, fluency) of translations at sentence level, at the level of full text, quality related to task at hand
 - ▶ Accuracy of translations (no missing info, no additional info)
 - ▶ What kind of errors does the system make, what kind of errors it doesn't make
 - ▶ How does it work with different genres of writing? (fine grained error analysis)
 - ▶ Is it possible to tag "bad translations" to be sent to human translators?
 - ▶ How much of post-editing will humans have with this output?
- ... and so on.

Going backwards with answers-4

- ▶ Q: If I don't have humans sitting there to do evaluations, how can I do automatic evaluations?
- ▶ A: Ask humans to translate some sentences, give those sentences to MT systems, and compare the degree of overlap between human and machine translations.
- ▶ There are different evaluation measures that focus on doing this.
- ▶ Advantage: automatic evaluation, so we don't have to conduct expensive, time-consuming evaluation studies anymore.
- ▶ Disadvantage?

Going backwards with answers-4

- ▶ Q: If I don't have humans sitting there to do evaluations, how can I do automatic evaluations?
- ▶ A: Ask humans to translate some sentences, give those sentences to MT systems, and compare the degree of overlap between human and machine translations.
- ▶ There are different evaluation measures that focus on doing this.
- ▶ Advantage: automatic evaluation, so we don't have to conduct expensive, time-consuming evaluation studies anymore.
- ▶ Disadvantage? - we don't know if the MT system is really good or just good on those sentences. We also don't know what is the best measure of closeness to human translation!

Going backwards with answers-5

- ▶ Q: Is doing human evaluations easy or difficult?
- ▶ A: Time and Cost. Reliability and Consistency. Meaningfully summarizing human judgements.
- ▶ Time and cost: can be addressed by doing crowdsourcing instead of looking for expert human translator judgements.
- ▶ Other issues: Difficult

Going backwards with answers-6

- ▶ Q: What is a good way to evaluate MT?
- ▶ A: Who is John Galt?

Why is it difficult to evaluate MT?

- ▶ Language variability - there is no single correct translation!
- ▶ Human evaluation is subjective
- ▶ For what purpose are we using MT?
- ▶ How do we evaluate evaluation?
- ▶ How do we decide how good is "good enough"?

Summary: MT evaluation is a research topic in itself! :)

Reference: Slides by Alon Lavie, MT Summit, 2011

State of MT today

- ▶ We have usable applications that work across many language pairs (Google Translate, Bing Translate, Babel, Facebook's translate etc)
- ▶ Not only are they used in different devices, they can also be deployed on your own websites, or applications as a service (if you know some programming, and have some money to license these)
e.g.: <https://azure.microsoft.com/en-us/pricing/details/cognitive-services/translator-text-api/>
e.g., <https://cloud.google.com/translate/pricing>
- ▶ They are still far from perfect, but are actively improving.
- ▶ One application that can have a potential impact on multiple aspects of human society in decades to come (something to think about for writing Q2 in Final Exam!)

Next Week

- ▶ Theme: Revision and some group exercises in the class
- ▶ Deadlines:
 - ▶ Final Exam, part 1: Writing a short summary of your Q1 topic (Due tomorrow)
 - ▶ Final Exam, part 2: Do a peer review for one person's Q1 draft above (IN THE CLASS, on 6th December. Don't miss the class!)
 - ▶ Final Exam, part 3: Submit Q1 and Q2, final versions in exams week (13th DEC)
- ▶ NOTE: No extensions on any of these deadlines!

Attendance Question

- ▶ Background:
 - ▶ There is this online competition called "Spooky Author Identification"
(<https://www.kaggle.com/c/spooky-author-identification>)
 - ▶ Task: you're challenged to predict the author of excerpts from horror stories by Edgar Allan Poe, Mary Shelley, and HP Lovecraft.
 - ▶ How does it work: First, they release some set of texts where authors are known, and are one of these.
 - ▶ You should then do "something" which can predict the author of a new text (who among these 3 is most likely?)
- ▶ Question: What ideas from this course are useful in solving this problem, conceptually speaking?