

# Introduction to Machine Translation

11-731 Machine Translation

Teruko Mitamura

Language Technologies Institute  
Carnegie Mellon University



## Outline

- Course Objectives
- Reading Materials
- Course Requirements
- Course Schedule and Topics
- Recommended Prerequisites
- Term Projects
- MT Conferences and Useful Information
- What is MT?



## Instructors

Will be taught jointly by:

- Teruko Mitamura (Course Leader)
- Robert Frederking
- Alon Lavie
- Guest Lecturer:
  - Jaime Carbonell, Eric Nyberg, Ralf Brown, Stephan Vogel



## Course Objectives

- Obtain a basic understanding of MT systems and MT-related issues
- Learn about theory and approaches in MT
- Learn about basic techniques for MT development
- Term Project: Obtain in-depth knowledge of one current topic in MT or Perform an analysis of a given MT problem (Research, Report and Presentation)



## Reading Materials

- Copies of some reading materials will be provided at Stacey Young's office (4503 NSH) soon
- Fee for each copy for compiling the materials
- Additional reading materials will be available from MT course web site or class handout, as necessary
- Additional resource lists: See the Syllabus



## Syllabus

<http://www.cs.cmu.edu/afs/cs/project/cmt-55/ti/Courses/731/www/syllabus-07.pdf>



## Course Requirements

- Homework: 3 assignments (25%)
  - Hands-on, learning by doing
- In Class Exam (closed book) (30%)
- Term Project (total 45%)
  - Class presentation and participation (10%)
  - Written Paper (35%)



## Course Topics

- Introduction to MT
  - Overview
  - History of MT
  - Principles of MT Research and Development
- Theory and Approaches for MT
  - Manual MT Evaluation and Human Factors
  - Automatic MT Evaluation
  - Transfer Methods
  - Interlingua MT
  - Statistical MT
  - Example-based MT
  - Multi-Engine MT



## Course Topics (2)

- Linguistic Analysis for MT System Development
  - Linguistics Knowledge Development
    - Analysis Knowledge Sources
    - Generation Knowledge Sources
  - Ambiguity & Ambiguity Resolution



## Course Topics (3)

- MT System Development: Case Study
  - Statistical MT
  - Example-based MT
  - Transfer MT: Learning MT Transfer Rules
  - Speech-to-Speech MT
  - Interlingua MT: KANTOO MT Systems
  - Business Case for MT and Commercial MT Systems



## Term Project

- Term Project is an individual piece of research, undertaken by the student with supervision by faculty.
- Project grades will be based on:
  - The quality of the research performed
  - In-class presentation and discussion
  - Final written report
- Term Project may synergize with preparation of an LTI thesis or a part of conference paper



## Term Project: Possible Topics

- An MT system design
- Study on a particular MT technique for:
  - Parsing
  - Generation
  - Ambiguity resolution
  - Controlled Language
  - Post-editing, etc.
- MT system evaluation
- Translation Aids: Interface Evaluation
- Comparative study of a set of current MT projects



## Examples of Past Projects

- Approaches for MT
  - Issues in Developing a Statistical MT System for Taiwanese and English
  - New directions in SMT: Linguistic structure in translation modeling and corpus-based decoding
  - A Sentence Length model for Statistical MT
  - A Limited Domain Statistical MT System for English to Hindi Translation



## Examples of Past Projects (2)

- Approaches for MT (cont.)
  - Applying New Clustering Algorithms for Generalization in EBMT
  - An Investigation of Corpus-based Approaches to Resolving Semantic Ambiguities
  - Improving the selection of the multi-engine MT system



## Examples of Past Projects (3)

- Translation in Cross-Lingual QA
  - Query Translation for Cross-Lingual Information Retrieval -- English to Chinese
  - Machine Translation of Sentences with Names Entities
  - Selecting Correct Translation of High-Density Language Pairs Using the WWW



## Examples of Past Projects (4)

- Various Topics
  - Developing Machine Translation Systems for Low Resource Languages
  - A Comparative Survey of Ambiguity Resolution in
  - Machine Translation Using Real-World Knowledge
  - The Significance of Part of Speech in Automatic Metrics for MT Evaluation



## Examples of Past Projects (5)

- Various Topics (2)
  - An Evaluation of the BLEU and NIST Metrics of Automatic Evaluation of MT (comparative analysis of proposed standards)
  - English-Portuguese Translation Using a Transfer-rule based system (a solution for a particular set of transfer problems)
  - Current Research Projects on Translating to Asian Languages (a survey of recent work)
  - Lattice Input for Speech-to-Speech MT (technical project on extending an existing system)



## Term Project Schedule

- Term Project Initial Proposal Due on Feb 14
  - The faculty will self-select the most appropriate advisor-to-project matches
- Feedback on the Initial Proposal
  - Students should meet with their project advisor at least once before the final proposal is submitted.
- Term Project Final Proposal Due on March 7
- Presentation Schedule: April 11 – May 2
  - About 45min (~30min. presentation + discussion)
- Final Term Project Paper Due: May 7



## Term Project Proposal

- Title of your project
- Description of project topic
- Outline of research to be performed
  - Goals (what you plan to achieve)
  - Plan (what tasks you will undertake)
  - Methodology (how you will carry out the work)
    - Specific software you would like to use (if applicable)
    - Specific data you would like to use (if applicable)
    - Research methods / strategies, etc.
  - Schedule of your research
  - References
- Propose a date for your class presentation including both 1<sup>st</sup> and 2<sup>nd</sup> choices.



## Term Project Evaluation Criteria

- The quality of the research performed
  - To show the depth of your knowledge with respect to the theme
  - Contribution to the MT field
  - The paper is sound technically and/or methodologically
  - Well-organized paper
  - Clearly written paper
  - Cover most relevant past work (references) in the topic



## Term Project Evaluation Criteria

- In-class presentation and in-class discussion (10%)
  - To show interests in other people's research topic
  - Practice asking questions and discussions in public
- Final written report (35%)
  - Usually 10-20 pages depending on the project (but not necessary longer the better; quality is more important)
  - Attach appendix if necessary
  - Make sure to include references
  - Personal conversation with advisors or other people should be mentioned in the paper and presentation if they contribute to the paper content in some way



## MT Conferences and Organizations

- International Conference on Theoretical and Methodological Issues in Machine Translation (TMI)
- MT Summit
- Conference of the Association for Machine Translation in the Americas (AMTA)
- The European Association for Machine Translation (EAMT)



## MT Conferences and Organizations

- Asia-Pacific Association for Machine Translation (AAMT)
- International Workshop on Controlled Language Applications (CLAW)
- See syllabus for more information



## Self Introduction

- Name
- Which topic(s) of MT are you interested in?
- Etc.



## What is MT?



## What is MT?

- “Computerized systems for translation from one natural language into another, with or without human assistance”
- Pre-editing:
  - Marking of input texts (e.g. proper nouns, phrase boundaries, disambiguation, etc.)
  - Controlled Language
  - Sub-language
- Post-editing:
  - MT output is revised by human translators



## What is MAHT (Machine-Aided Human Translation)?



## What is MAHT?

- Machine-Aided Human Translation or Computer-Aided (or Assisted) Translation
  - Computer-based translation tools which support translators by providing access to dictionaries and terminology and texts databases, interacting with word processing, text editing or printing.
  - e.g., Translation Memory Systems



## Popular Conceptions and Misconceptions

<http://www.essex.ac.uk/linguistics/clmt/MTbook/HTML/book.html>

- “MT is a waste of time because you will never make a machine that can translate Shakespeare.”
- “Generally, the quality of translation you can get from an MT system is very low. This makes them useless in practice.”
- “MT threatens the jobs of translators.”



## Popular Conceptions and Misconceptions (2)

- “MT systems are machines and buying an MT system should be very much like buying a car.”
- “In practice, before an MT system becomes really useful, a user will typically have to invest a considerable amount of effort in customizing it.”



Questions?

