Machine Translation Overview

Alon Lavie
Language Technologies Institute
Carnegie Mellon University

LTI Immigration Course
August 22, 2011

Machine Translation: History

- **1946:** MT is one of the first conceived applications of modern computers (A.D. Booth, Alan Turing)
- **1954:** The “Georgetown Experiment” Promising “toy” demonstrations of Russian-English MT
- **Late 1950s and early 1960s:** MT fails to scale up to “real” systems
- **1966:** ALPAC Report: MT recognized as an extremely difficult, “AI-complete” problem. Funding disappears
- **1968:** SYSTRAN founded
- **1985:** CMU “Center for Machine Translation” (CMT) founded
- **Late 1980s and early 1990s:** Field dominated by rule-based approaches – KBMT, KANT, Eurotra, etc.
- **1992:** “Noisy Channel” Statistical MT models invented by IBM researchers (Brown, Della Pietra, et al.). CANDIDE
- **Mid 1990s:** First major DARPA MT Program. PANGLOSS
- **Late 1990s:** Major Speech-to-Speech MT demonstrations: C-STAR
- **1999:** JHU Summer Workshop results in GIZA
- **2000s:** Large DARPA Funding Programs – TIDES and GALE
- **2003:** Och et al introduce Phrase-based SMT. PHARAOH
- **2006:** Google Translate is launched
- **2007:** Koehn et al release MOSES

Machine Translation: Where are we today?

- Age of Internet and Globalization – great demand for translation services and MT:
  - Multiple official languages of UN, EU, Canada, etc.
  - Software Localization and documentation dissemination for large manufacturers (Microsoft, Intel, Apple, EBay, ALCOA, etc.)
  - Language and translation services business sector estimated at $26 Billion worldwide in 2010 and growing at a healthy pace
  - Volume of online content growing exponentially
- Economic incentive is still primarily within a small number of language pairs
- Some fairly decent commercial products in the market for these language pairs
  - Product of rule-based systems after many years of development: SYSTRAN, PROMT, others...
  - New generation of data-driven "statistical" MT systems: SDL/Language Weaver, Asia Online, others...
- Web-based (mostly free) MT services: Google, MS-Bing, Babelfish, others...
- Pervasive MT between many language pairs still non-existent, but some significant progress in recent years

How Does MT Work?

- All modern MT approaches are based on building translations for complete sentences by putting together smaller pieces of translation
- Core Questions:
  - What are these smaller pieces of translation? Where do they come from?
  - How does MT put these pieces together?
  - How does the MT system pick the correct (or best) translation among many options?

Core Challenges of MT

- **Ambiguity and Language Divergences:**
  - Human languages are highly ambiguous, and differently in different languages
  - Ambiguity at all “levels”: lexical, syntactic, semantic, language-specific constructions and idioms
- **Amount of required knowledge:**
  - Translation equivalencies for vast vocabularies (several 100k words and phrases)
  - Syntactic knowledge (how to map syntax of one language to another), plus more complex language divergences (semantic differences, constructions and idioms, etc.)
  - **How** do you acquire and construct a knowledge base that big that is (even mostly) correct and consistent?

How to Tackle the Core Challenges

- **Manual Labor:** 1000s of person-years of human experts developing large word and phrase translation lexicons and translation rules.

  Example: Systran's RBMT systems.
- **Lots of Parallel Data:** data-driven approaches for finding word and phrase correspondences automatically from large amounts of sentence-aligned parallel texts. Example: Statistical MT systems.
- **Learning Approaches:** learn translation rules automatically from small amounts of human translated and word-aligned data. Example: AVENUE's Statistical XFER approach.
- **Simplify the Problem:** build systems that are limited-domain or constrained in other ways. Examples: CATALYST, NESPOLE!.

Rule-based vs. Data-driven Approaches to MT

- What are the pieces of translation? Where do they come from?
  - Rule-based: large-scale “clean” word translation lexicons, manually constructed over time by experts
  - Data-driven: broad-coverage word and multi-word translation lexicons, learned automatically from available sentence-parallel corpora
- How does MT put these pieces together?
  - Rule-based: large collections of rules, manually developed over time by human experts, that map structures from the source to the target language
  - Data-driven: a computer algorithm that explores millions of possible ways of putting the small pieces together, looking for the translation that statistically looks best

Rule-based vs. Data-driven Approaches to MT

- How does the MT system pick the correct (or best) translation among many options?
  - Rule-based: Human experts encode preferences among the rules designed to prefer creation of better translations
  - Data-driven: a variety of fitness and preference scores, many of which can be learned from available training data, are used to model a total score for each of the millions of possible translation candidates; algorithm then selects and outputs the best scoring translation

Rule-based vs. Data-driven Approaches to MT

- Why have the data-driven approaches become so popular?
  - We can now do this!
    - Increasing amounts of sentence-parallel data are constantly being created on the web
    - Advances in machine learning algorithms
    - Computational power of today's computers can train systems on these massive amounts of data and can perform these massive search-based translation computations when translating new texts
  - Building and maintaining rule-based systems is too difficult, expensive and time-consuming
  - In many scenarios, it actually works better!

Statistical MT (SMT)

- Data-driven, most dominant approach in current MT research
- Originally proposed by IBM in early 1990s: a direct, purely statistical, model for MT
- Evolved from word-level translation to phrase-based translation
- Main Ideas:
  - Training: statistical "models" of word and phrase translation equivalence are learned automatically from bilingual parallel sentences, creating a bilingual "database" of translations
  - Decoding: new sentences are translated by a program (the decoder), which matches the source words and phrases with the database of translations, and searches the "space" of all possible translation combinations.

Statistical MT: Major Challenges

- Current approaches are too naïve and “direct”:
  - Good at learning word-to-word and phrase-to-phrase correspondences from data
  - Not good enough at learning how to combine these pieces and reorder them properly during translation
  - Learning general rules requires much more complicated algorithms and computer processing of the data
  - The space of translations that is “searched” often doesn’t contain a perfect translation
  - The fitness scores that are used aren’t good enough to always assign better scores to the better translations → we don’t always find the best translation even when it’s there!
  - MERT is brittle, problematic and metric-dependent!
- Solutions:
  - Google solution: more and more data!
  - Research solution: “smarter” algorithms and learning methods

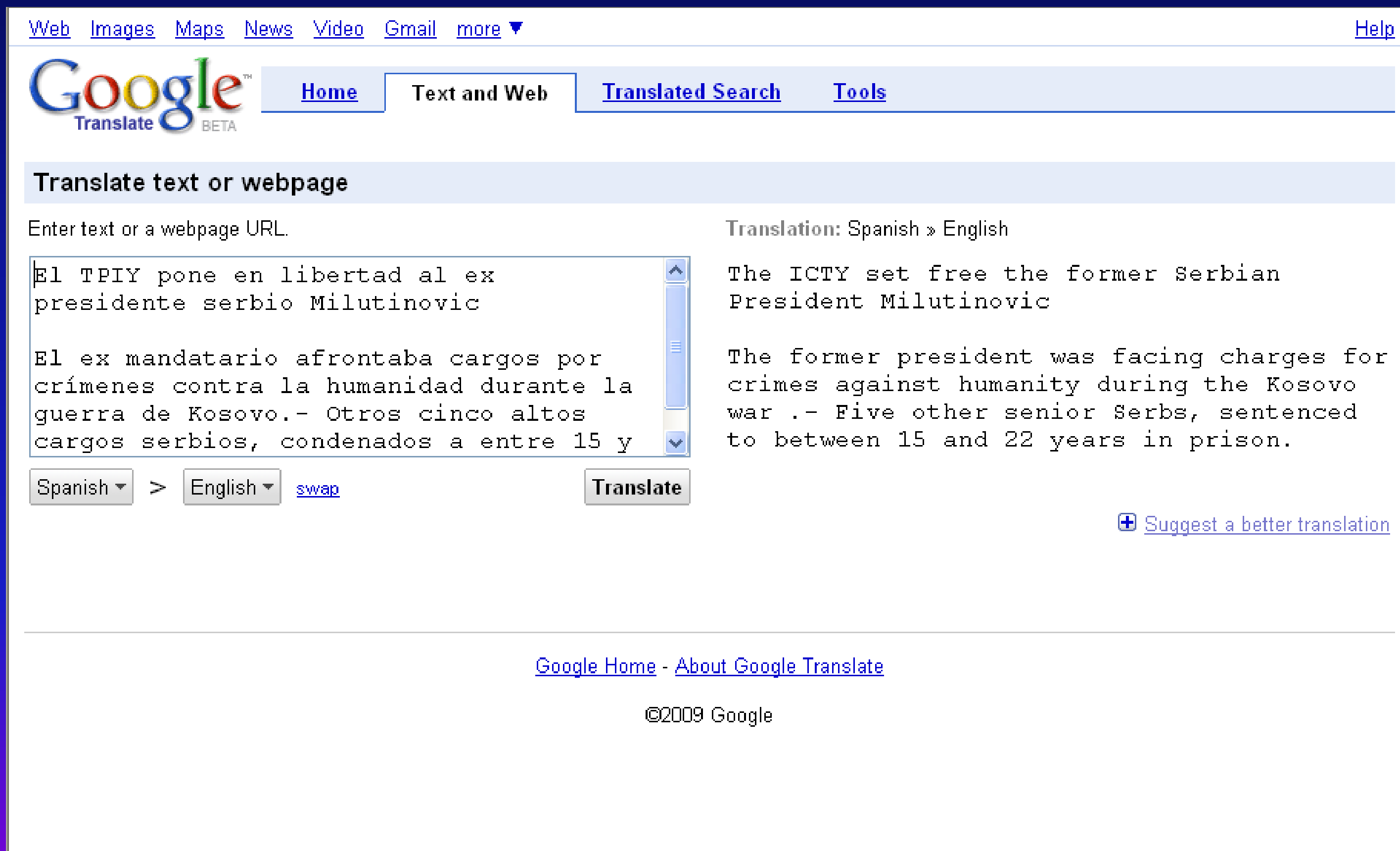
Rule-based vs. Data-driven MT

Rule-based	Data-driven
We thank all participants of the whole world for their comical and creative drawings; to choose the victors was not easy task! Click here to see work of winning European of these two months, and use it to look at what the winning of USA sent us.	We thank all the participants from around the world for their designs cocasses and creative; selecting winners was not easy! Click here to see the artwork of winners European of these two months, and disclosure to look at what the winners of the US have been sending.

Representative Example: Google Translate

- http://translate.google.com

Google Translate



Google Translate

Types of MT Applications:

- **Assimilation:** multiple source languages, uncontrolled style/topic. General purpose MT, no semantic analysis. (GP FA or GP HQ)
- **Dissemination:** one source language, controlled style, single topic/domain. Special purpose MT, full semantic analysis. (FA HQ)
- **Communication:** Lower quality may be okay, but system robustness, real-time required.

Approaches to MT: Vaquois MT Triangle

Interlingua

Give-information+personal-data (name=alon_lavie)

Analysis

Generation

Transfer

[$_{s}$ [$_{vp}$ accusative_pronoun "chiamare" proper_name]]

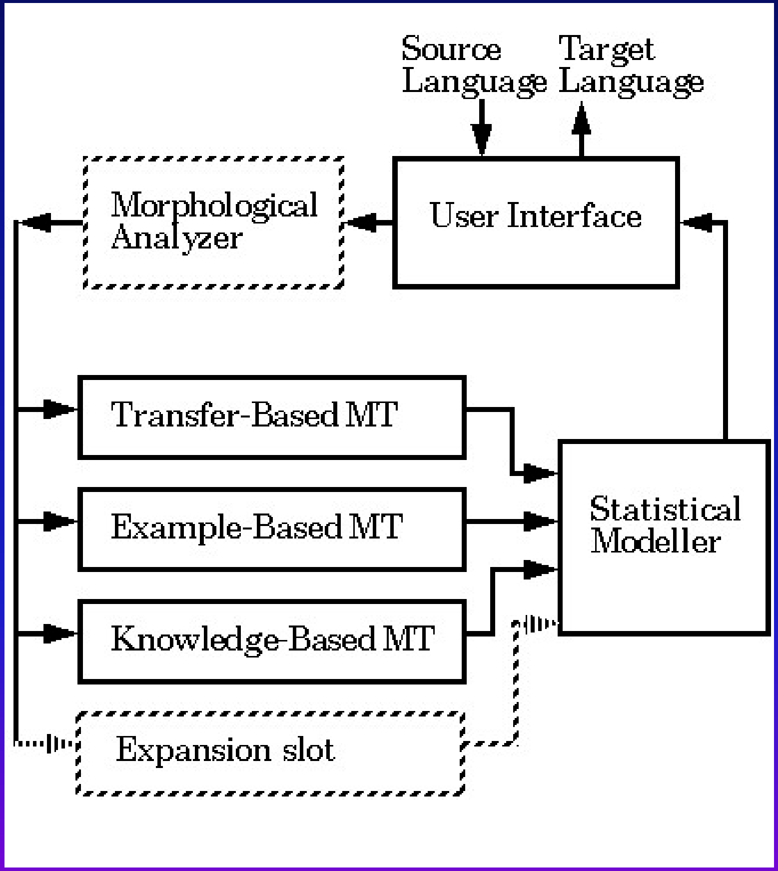
[$_{s}$ [$_{np}$ [possessive_pronoun "name"]] [$_{vp}$ "be" proper_name]]

Direct

Mi chiamo Alon Lavie

My name is Alon Lavie

Multi-Engine MT



- Apply several MT engines to each input in parallel
- Create a combined translation from the individual translations
- Goal is to combine strengths, and avoid weaknesses.
- Along all dimensions: domain limits, quality, development time/cost, run-time speed, etc.
- Various approaches to the problem

Speech-to-Speech MT

- Speech just makes MT (much) more difficult:
  - Spoken language is messier
    - False starts, filled pauses, repetitions, out-of-vocabulary words
    - Lack of punctuation and explicit sentence boundaries
  - Current Speech technology is far from perfect
- Need for speech recognition and synthesis in foreign languages
- Robustness: MT quality degradation should be proportional to SR quality
- Tight Integration: rather than separate sequential tasks, can SR + MT be integrated in ways that improves end-to-end performance?

MT at the LTI

- LTI originated as the Center for Machine Translation (CMT) in 1985
- MT continues to be a prominent sub-discipline of research with the LTI
- Active research on all main approaches to MT
- Leader in the area of speech-to-speech MT
- Multi-Engine MT (MEMT)
- MT Evaluation (METEOR)
- Spin-off Companies:
  - Jibbigo (speech translation on mobile devices)
  - Safaba (MT solutions for enterprises and LSPs)

MT Faculty at LTI

- Alon Lavie
- Stephan Vogel
- Ralf Brown
- Jaime Carbonell
- Lori Levin
- Noah Smith
- Alan Black
- Florian Metze
- Alex Waibel
- Teruko Mitamura
- Eric Nyberg

Summary

- Main challenges for current state-of-the-art MT approaches - Coverage and Accuracy:
  - Acquiring broad-coverage high-accuracy translation lexicons (for words and phrases)
  - learning structural mappings between languages from parallel word-aligned data
  - overcoming syntax-to-semantics differences and dealing with constructions
  - Stronger Target Language Modeling
  - Context-dependent modeling and adaptation
  - Novel algorithms for model acquisition and decoding

Questions...