Machine Translation (MT)

Translating text from one language to another.

Machine Translation

- Translating text from one language to another is a task challenging even for humans to try to fully capture the style and nuanced meaning of the original
- While research focuses on trying to produce the fully-automatic, high-quality translation, there are many tasks for which a rough translation is sufficient
- The differences between languages include systematic differences that can be modeled in some way and idiosyncratic and lexical differences that must be dealt with one by one.
- Machine translation focuses on
  - Faithfulness – the meaning of the text has been preserved
  - Fluency – the translated text sounds natural to a native speaker

Why MT is hard

- Given the Japanese phrase
  fukaku hansei shite orimasu
- If this is translated to English as
  we apologize
  it is not faithful to the original meaning
- But if we translate it as
  we are deeply reflecting (on our past behavior, and what we did wrong, and how to avoid the problem next time)

  the translation is not fluent.

  Example from Jurafsky and Martin text.

Differences between languages

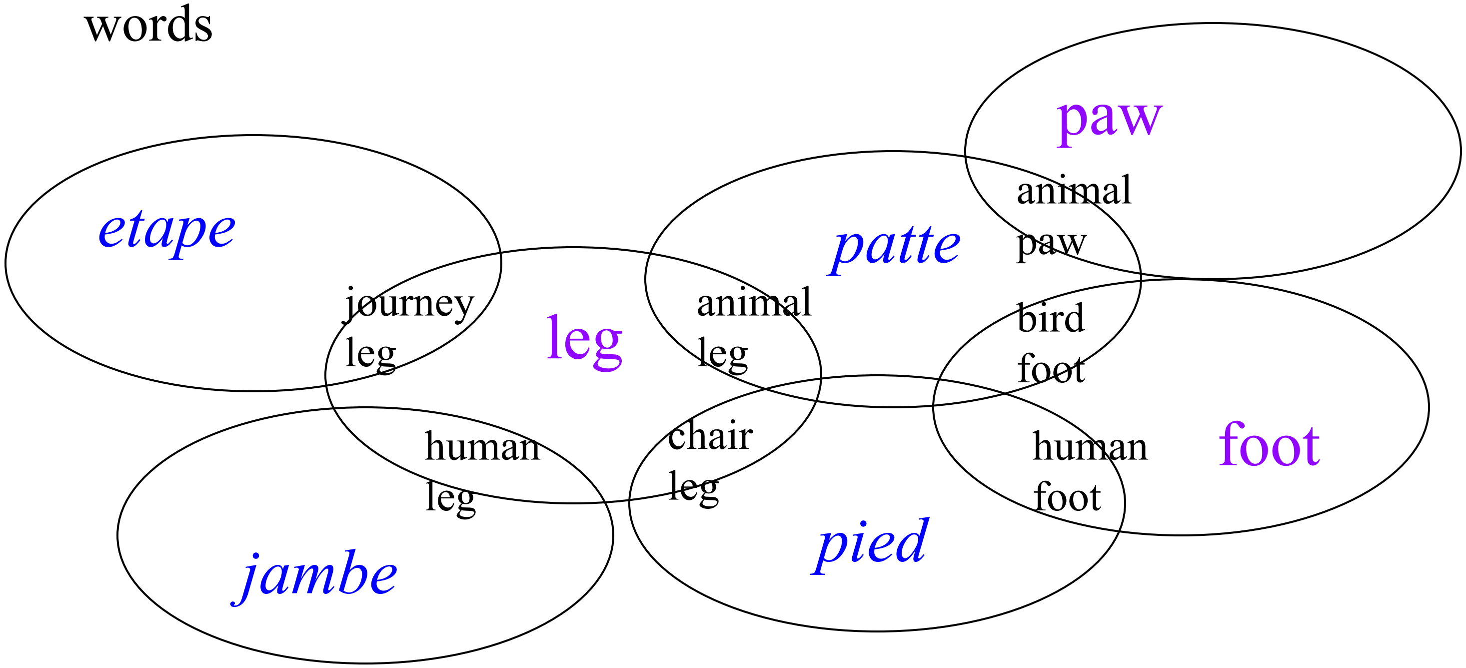
- Morphological differences:
  - Number of morphemes per word and sentence
    - Isolating languages: Vietnamese and Cantonese, each word has one morpheme
    - Polysynthetic languages: "Eskimo", a single word has many morphemes corresponding to a complete sentence.
  - Degree to which morphemes are segmentable
    - Agglutinative, morphemes have clean boundaries (Turkish)
    - Fusion languages, single affix may have multiple morphemes (Russian)

Differences between languages

- Syntactic differences
  - Basic word order of verbs, subjects and objects
    - SVO: English, Mandarin, French, German, …
    - SOV: Hindi, Japanese
    - VSO: Classical Arabic and Biblical Hebrew
  - Head marking and dependent marking languages
    - Mark relation between dependent and head on the head
      - English marks possessive on dependent: *the man's house*
      - Hungarian marks possessive on the head noun: (Hungarian equivalent of:) *the man house-his*
  - Direction of motion with respect to verb
    - English direction on particle: *the bottle floated out*
    - Spanish direction on verb: *la botella salio' flotando*
  - Grammatical constraints on matching gender-marked words
  - Many others . . .

Differences between languages

- Semantic differences
  - Lexical gap
    - One language doesn't have a word for concept in another
  - Differences in way that conceptual space is divided up for different words



The complex overlap between English leg, foot, etc. and various French translations. (Jurafsky & Martin, Figure 21.2)

Classical MT/Machine Translation

- In this line of MT research, approaches can be classified according to the level of unit of translation
  - Utilizes word translation dictionaries
  - Direct translation uses a word translation approach
  - Transfer approaches use syntactic phrase and semantic units as the unit of translation

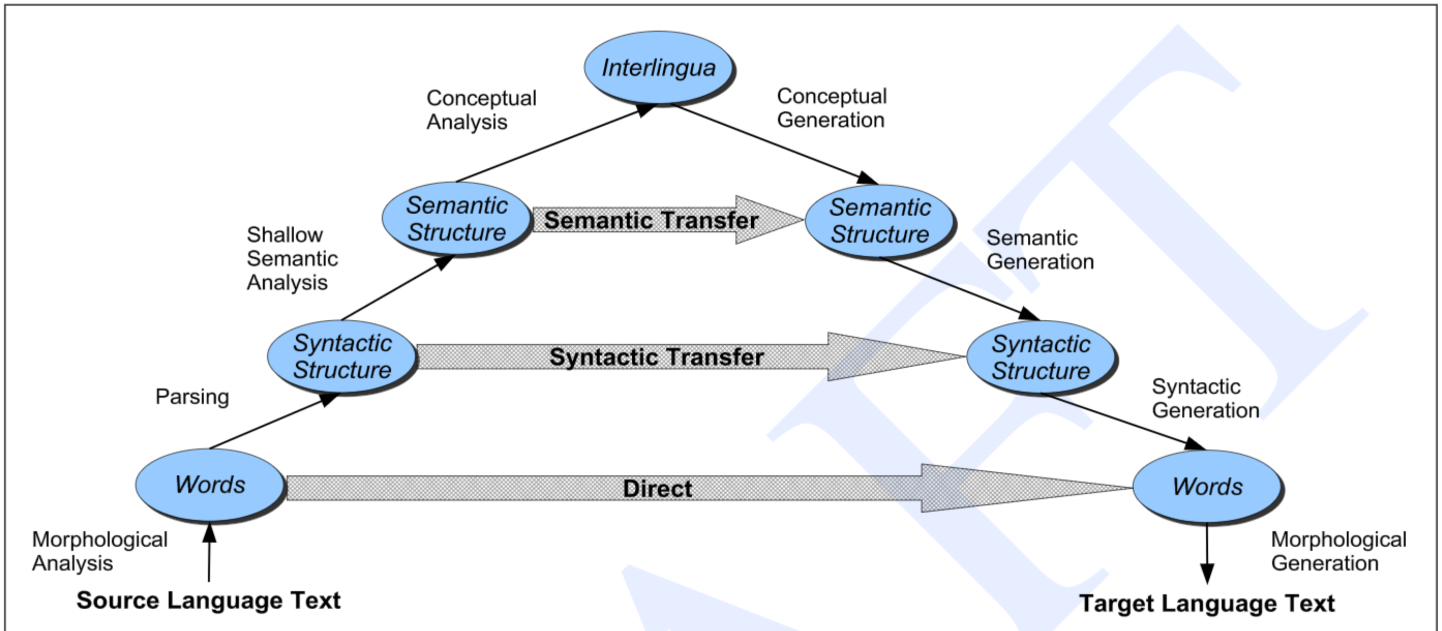


Figure 25.3 The Vauquois triangle.

Statistical Approaches

- Build probabilistic models of faithfulness and fluency and combine the models to get the most probable translation.
- Modeled as a noisy channel "pretend that the foreign input F is a corrupted version of the target language output E and the task is to discover the hidden sentence E that generated the observed sentence F."
  - Informally, we refer to translating from French to English
- Requires two models
  - Language model to compute P(E), probability that any sequence E of English words is a sentence (fluency)
  - Translation model to compute P(F|E), conditional probability that French sentence F was a translation of an English sentence E (faithfulness)
- Given French sentence f, its translation e is

  arg max (all e in E) P(e) * P(f | e)

  - Note that this appears backwards to translate from French to English, but we invoke Bayes theorem to define the decoder.

Statistical Language Models

- Language model to compute P(E)
  - In practice, learn probabilities of bigrams in the language to be translated from instead of entire sentences
  - Translation has improved greatly due to large corpora
    - See Google Translate
- Translation model to compute P(F|E)
  - Learn probabilities from parallel corpora
  - Model the translation as word translation combined with alignment prob.
    - E: *And the program has been implemented.*
    - F: *Le programme a ete mis en application*.
    - Alignment variables: (2, 3, 4, 5, 6, 6, 6) gives

Le	*-> the*	*mis*	*-> implemented*
Programme	*-> program*	*en*	*-> implemented*
a	*-> has*	*application*	*-> implemented*
ete	*-> been*		

Alignment and Parallel Corpora

- The translation model uses probabilities of word alignment
- Word alignment models are automatically trained from parallel corpora where each document is given in two or more languages
  - Hansard Corpus
    - Canadian parliament documents for French, English and a variety of native American languages
  - United Nations proceedings documents
  - LDC (Linguistic Data Consortium) has corpora in several language pairs
- Literary parallel corpora are not as suitable because of the stronger presence of literary devices, such as metaphor

MT Evaluation

- Human raters can evaluate along the two dimensions of fluency and fidelity (and there are several individual metrics for each of these dimensions)
- BLEU automatic evaluation system
  - Evaluation corpus contains human generated translations
  - Metrics evaluate how closely the system-generated translations correspond to the human ones