Machine Translation

J&M’s Chapter 25
Machine Translation

• Automatically translate one natural language into another.

Mary didn’t slap the green witch.

Maria no dió una bofetada a la bruja verde.

Mary not give a slap to the witch green
Ambiguity Resolution is Required for Translation

• Syntactic and semantic ambiguities must be properly resolved for correct translation:
  – “John plays the guitar.” → “John gitar minavazad.”
  – “John plays soccer.” → “John football bazi mikonad.”

• An early MT system gave the following results when translating from English to Russian and then back to English:
  – “The spirit is willing but the flesh is weak.” ⇒ “The liquor is good but the meat is spoiled.”
  – “Out of sight, out of mind.” ⇒ “Invisible idiot.”
Word Alignment

- Shows mapping between words in one language and the other.

Mary didn’t slap the green witch.

Maria no dio una bofetada a la bruja verde.

Mary not give a slap to the witch green
Translation Quality

- Achieving literary quality translation is very difficult.
- Existing MT systems can generate rough translations that frequently at least convey the gist of a document.
- High quality translations possible when specialized to narrow domains, e.g., weather forecasts.
- Some MT systems used in computer-aided translation in which a bilingual human post-edits the output to produce more readable accurate translations.
- Frequently used to aid localization of software interfaces and documentation to adapt them to other languages.
Linguistic Issues Making MT Difficult

- Morphological issues with *isolating* (e.g., Vietnamese) vs. *polysynthetic* (e.g., Eskimo), and *agglutinative* (e.g., Turkish) vs. *fusion* (e.g., Russian) languages.

- Syntactic variation between *SVO* (e.g., English), *SOV* (e.g., Persian), and *VSO* (e.g., Arabic) languages.
  - SVO languages use prepositions
  - SOV languages use postpositions

- *Pro-drop* languages regularly omit subjects that must be inferred.
Lexical Gaps

• Some words in one language do not have a corresponding term in the other.
  • Rivière (river that flows into ocean) and fleuve (river that does not flow into ocean) in French
  • Schedenfraude (feeling good about another’s pain) in German.
  • Oyakoko (filial piety: devoting to parents) in Japanese
Lexical Gaps

Relation between English leg, foot, and paw, to the French jambe, pied, patte, and etape
Vauquois Triangle

Conceptual Analysis → Interlingua

Semantic Transfer → Semantic structure

Syntactic Transfer → Syntactic structure

Direct translation → Words

Source Language

Morphological Analysis

Target Language

Morphological Generation
Direct Transfer

- **Morphological Analysis**
  - Mary didn’t slap the green witch. →
    Mary DO-PAST not slap the green witch.

- **Lexical Transfer**
  - Mary DO-PAST not slap the green witch.
    - Maria PAST no dar una bofetada a la verde bruja.

- **Lexical Reordering**
  - Maria no dar: PAST una bofetada a la bruja verde.

- **Morphological generation**
  - Maria no dió una bofetada a la bruja verde.
Syntactic Transfer

• Simple lexical reordering does not adequately handle more dramatic reordering such as that required to translate from an SVO to an SOV language.

• Need syntactic transfer rules that map parse tree for one language into one for another.
  – English to Spanish:
    • NP → Adj Noun ⇒ NP → Noun Adj
  – English to Japanese:
    • VP → V NP ⇒ VP → NP V
    • PP → P NP ⇒ PP → NP P
Syntactic transformation from English (SVO) to Japanese (SOV)
A further sketch of the transfer approach
Semantic Transfer

• Some transfer requires semantic information.
• Semantic roles can determine how to properly express information in another language.
• In Chinese, PPs that express a goal, destination, or benefactor occur before the verb but those expressing a recipient occur after the verb.
• Transfer Rule
  – English to Chinese
    • $\text{VP} \rightarrow \text{V PP[+Goal]} \Rightarrow \text{VP} \rightarrow \text{PP[+Goal]} \text{ V}$
Chinese Goal PPs often occur preverbally.

Jackie Chan went to Hong Kong.
Mary did not slap the green witch.

Interlingua representation of

EVENT: SLAPPING
AGENT: Mary
TENSE: PAST
POLARITY: NEGATIVE
THEME: [WITCH [DEFINITEDEF [ATTRIBUTES [HAS-COLOR GREEN]]]]
Statistical MT

• Manually encoding comprehensive bilingual lexicons and transfer rules is difficult.
• SMT acquires knowledge needed for translation from a parallel corpus that contains the same set of documents in two languages.
• The Canadian parliamentary proceedings in French and English is a well-known parallel corpus.
• First align the sentences in the corpus based on simple methods.
Picking a Good Translation

• A good translation should be \textit{faithful} and correctly convey the information and tone of the original source sentence.

• A good translation should also be \textit{fluent}, grammatically well structured and readable in the target language.

• Final objective:

\[
T_{\text{best}} = \arg\max_{T \in \text{Target}} \text{faithfulne ss}(T, S) \text{ fluency}(T)
\]


Noisy Channel Model

**GENERATIVE DIRECTION**

“CHANNEL SOURCE E”

P(E) → E: Mary did not slap the green witch

“CHANNEL OUTPUT F”

P(F|E) → F: María no dio una bofetada a la bruja verde

**BEST TARGET LANGUAGE SENTENCE E**

**ARGMAX**

**LANGUAGE MODEL P(E)**

**TRANSLATION MODEL P(F|E)**

**DECODING DIRECTION**
Bayesian Analysis of Noisy Channel

\[ \hat{E} = \text{argmax}_{E \in \text{English}} P(E \mid F) \]
\[ = \text{argmax}_{E \in \text{English}} \frac{P(F \mid E)P(E)}{P(F)} \]
\[ = \text{argmax}_{E \in \text{English}} P(F \mid E)P(E) \]

Translation Model    Language Model

A decoder determines the most probable translation \( \hat{E} \) given \( F \)
Language Model

- Use a standard \textit{n}-gram language model for \( P(E) \).
- Can be trained on a large, unsupervised mono-lingual corpus for the target language \( E \).
- Could use a more sophisticated PCFG language model to capture long-distance dependencies.
Translation Probabilities

• Assuming a *phrase aligned* parallel corpus is available or constructed that shows matching between phrases in $E$ and $F$.

• Then compute (MLE) estimate of $\phi$ based on simple frequency counts.

$$\phi(\bar{f}, \bar{e}) = \frac{\text{count}(\bar{f}, \bar{e})}{\sum_{\tilde{f}} \text{count}(\tilde{f}, \bar{e})}$$
Evaluating MT

- Human subjective evaluation is the best but is time-consuming and expensive.
- Automated evaluation comparing the output to multiple human reference translations is cheaper and correlates with human judgments.
Computer-Aided Translation Evaluation

• **Edit cost**: Measure the number of changes that a human translator must make to correct the MT output.
  
  – Number of words changed
  
  – Amount of time taken to edit
  
  – Number of keystrokes needed to edit
Automatic Evaluation of MT

• Collect one or more human *reference translations* of the source.
• Compare MT output to these reference translations.
• Score result based on similarity to the reference translations.
  – BLEU
  – NIST
  – TER
  – METEOR
BLEU

• Determine number of $n$-grams of various sizes that the MT output shares with the reference translations.
• Compute a modified precision measure of the $n$-grams in MT result.
**BLEU Example**

Cand 1: Mary no slap the witch green
Cand 2: Mary did not give a smack to a green witch.

Ref 1: Mary did not slap the green witch.
Ref 2: Mary did not smack the green witch.
Ref 3: Mary did not hit a green sorceress.

**Cand 1 Unigram Precision: 5/6**
Cand 1: Mary no slap the witch green.  
Cand 2: Mary did not give a smack to a green witch.

Ref 1: Mary did not slap the green witch.  
Ref 2: Mary did not smack the green witch.  
Ref 3: Mary did not hit a green sorceress.

Cand 1 Bigram Precision: 1/5
BLEU Example

Cand 1: Mary no slap the witch green.
Cand 2: Mary did not give a smack to a green witch.

Ref 1: Mary did not slap the green witch.
Ref 2: Mary did not smack the green witch.
Ref 3: Mary did not hit a green sorceress.

Clip match count of each $n$-gram to maximum count of the $n$-gram in any single reference translation

Cand 2 Unigram Precision: 7/10
BLEU Example

Cand 1: Mary no slap the witch green.
Cand 2: Mary did not give a smack to a green witch.

Ref 1: Mary did not slap the green witch.
Ref 2: Mary did not smack the green witch.
Ref 3: Mary did not hit a green sorceress.

Cand 2 Bigram Precision: 4/9
**Modified N-Gram Precision**

- Average $n$-gram precision over all $n$-grams up to size $N$ (typically 4) using geometric mean.

\[
p_n = \frac{\sum_{C \in \text{corpus}} \sum_{n-gram \in C} \text{count}_{\text{clip}}(n-\text{gram})}{\sum_{C \in \text{corpus}} \sum_{n-gram \in C} \text{count}(n-\text{gram})}
\]

\[
p = \sqrt[4]{\prod_{n=1}^{N} p_n}
\]

**Cand 1:**  
\[
p = 2\sqrt[4]{\frac{5}{6} \cdot \frac{1}{5}} = 0.408
\]

**Cand 2:**  
\[
p = 2\sqrt[4]{\frac{7}{10} \cdot \frac{4}{9}} = 0.558
\]
Brevity Penalty

- Not easy to compute recall to complement precision since there are multiple alternative gold-standard references and don’t need to match all of them.
- Instead, use a penalty for translations that are shorter than the reference translations.
- Define effective reference length, $r$, for each sentence as the length of the reference sentence with the largest number of $n$-gram matches. Let $c$ be the candidate sentence length.

$$BP = \begin{cases} 
1 & \text{if } c > r \\
\frac{e^{(1-r/c)}}{e^{1-c}} & \text{if } c \leq r 
\end{cases}$$
BLEU Score

• Final BLEU Score: \(\text{BLEU} = BP \times \rho\)
  
  Cand 1: Mary no slap the witch green.
  
  Best Ref: Mary did not slap the green witch.

  \[
  c = 6, \quad r = 7, \quad BP = e^{(1-7/6)} = 0.846 \\
  \text{BLEU} = 0.846 \times 0.408 = 0.345
  \]

  Cand 2: Mary did not give a smack to a green witch.
  
  Best Ref: Mary did not smack the green witch.

  \[
  c = 10, \quad r = 7, \quad BP = 1 \\
  \text{BLEU} = 1 \times 0.558 = 0.558
  \]