40417: Artificial Intelligence

Machine Translation (Reading: J+M Ch10)

Instructor: Gholamreza Ghassem-Sani --- Sharif University of Technology

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

- Syntactic and semantic ambiguities must be properly resolved for correct translation:
  - "John plays the guitar." → "John gitar minavazad."
  - "John plays soccer." → "John footbal 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 dió una bofetada a la bruja verde.

Mary not give a slap to the witch green

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

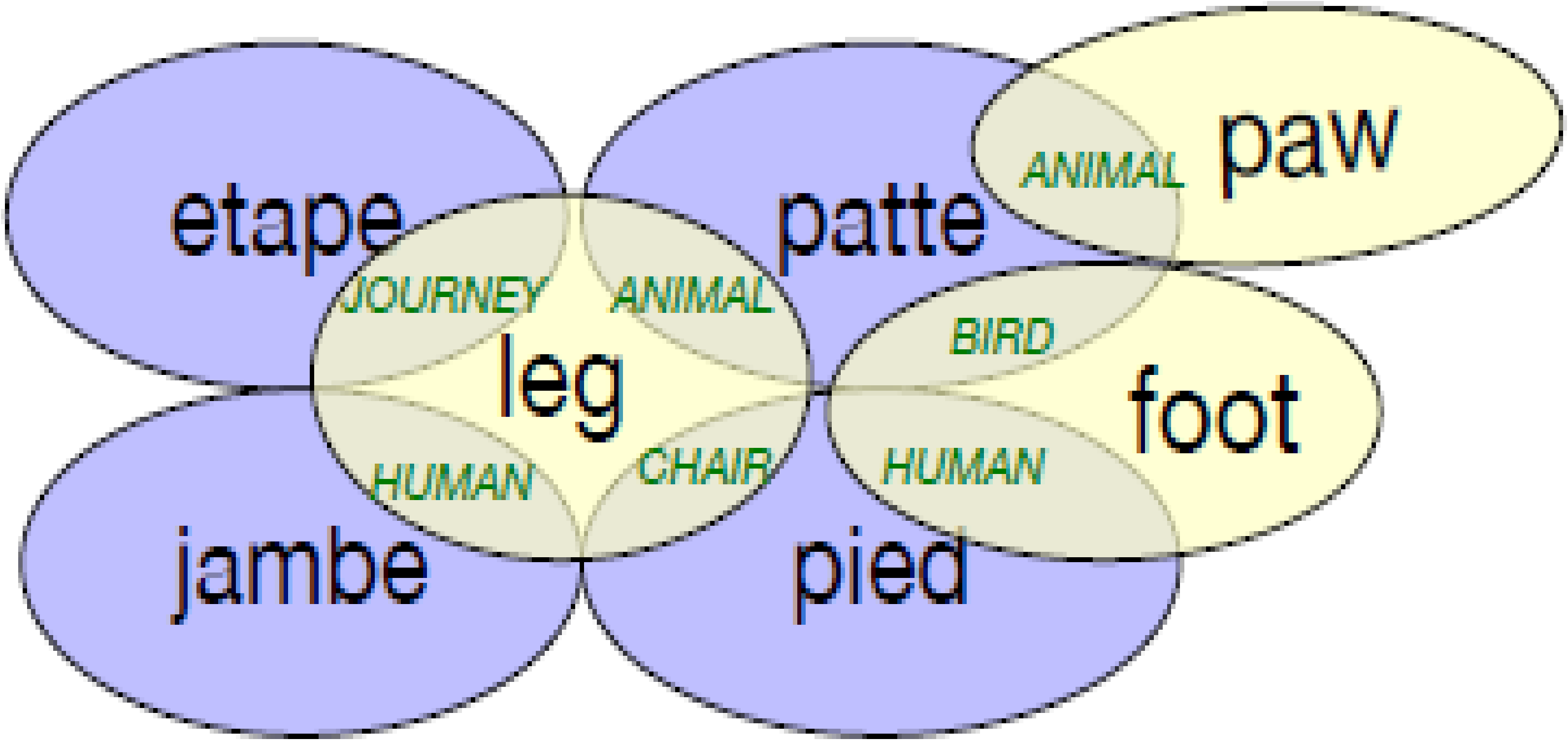
- ***Pro-drop*** languages regularly omit subjects that must be inferred.

Lexical Gaps

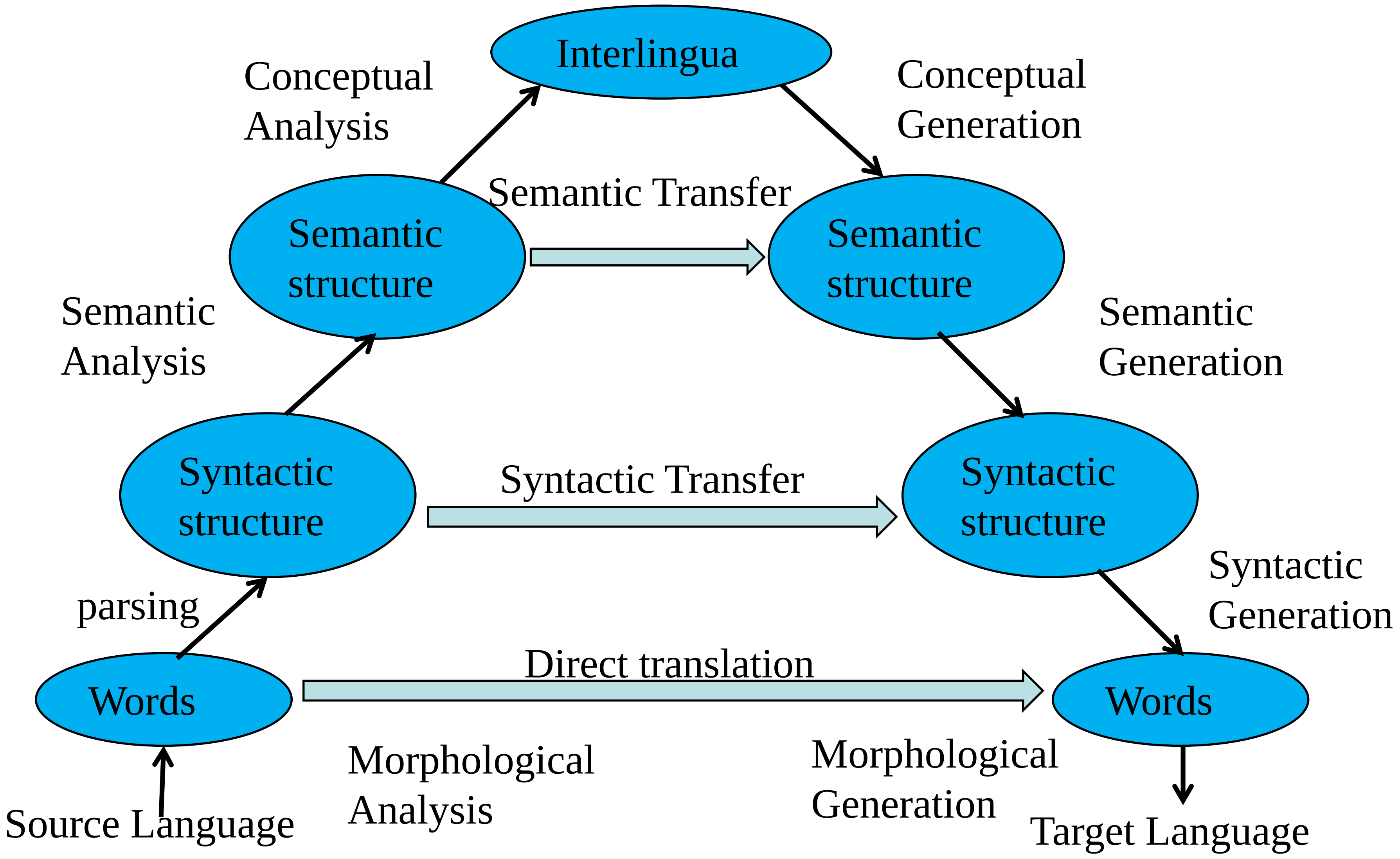
- Some words in one language do not have a corresponding term in the other.
  - Rivière (river that flows into ocean) and
    fleuve (river that does not flow into ocean) in French
  - Schedenfraude (feeling good about another's pain) in German
  - Oyakoko (filial piety: devoting to parents) in Japanese

Lexical Gaps

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



Vauquois Triangle

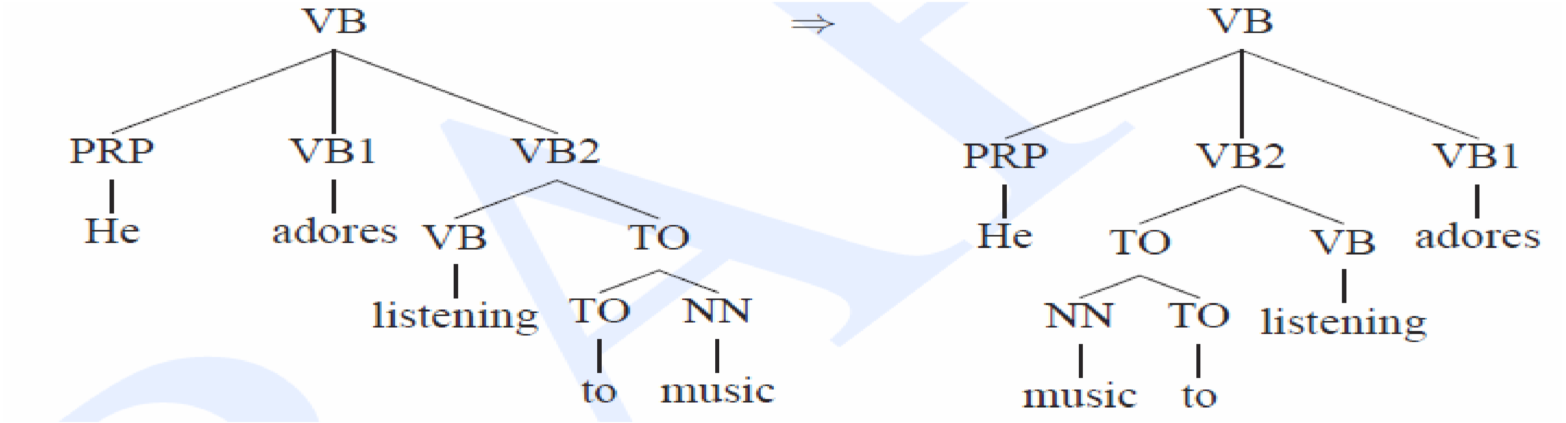


Direct Transfer

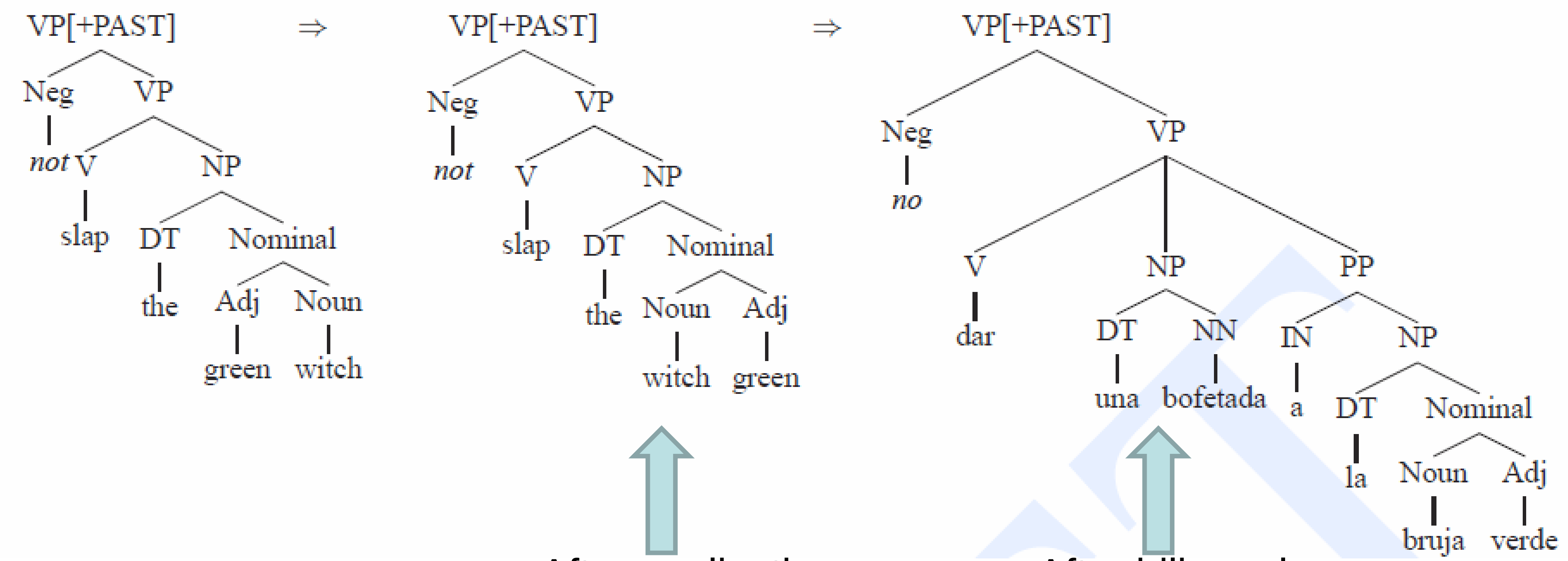
- Morphological Analysis
  - Mary didn’t slap the green witch. →
    Mary DO-PAST not slap the green witch.
- Lexical Transfer
  - Mary DO-PAST not slap the green witch.
  - Maria PAST no dar una bofetada a la verde bruja.
- Lexical Reordering
  - Maria no dar:PAST una bofetada a la bruja verde.
- Morphological generation
  - Maria no dió una bofetada a la bruja verde.

Syntactic Transfer

- Dealing with word order variation (e.g. English SVO → Japanese SOV)
- Example transfer rules:
  - VP → V NP := VP → NP V
  - PP → P NP := PP → NP P



A further sketch of the transfer approach



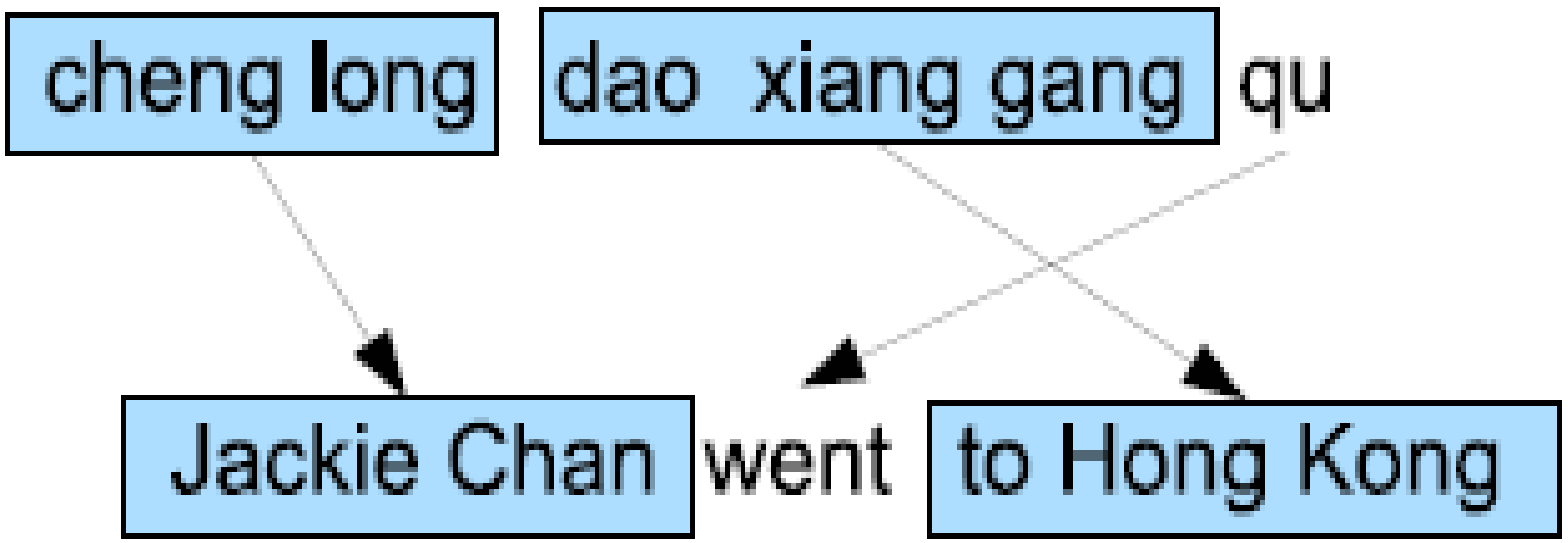
After application
of transfer rules

After bilingual
dictionary lookup

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
    - VP → V PP[+Goal] ⇒ VP → PP[+Goal] V

Semantic Transfer



Interlingua representation

$$
\begin{bmatrix}
\text{EVENT} & \text{SLAPPING} \\
\text{AGENT} & \text{MARY} \\
\text{TENSE} & \text{PAST} \\
\text{POLARITY} & \text{NEGATIVE} \\
\text{THEME} & \begin{bmatrix} \text{WITCH} & \\ \text{DEFINITENESS} & \text{DEF} \\ \text{ATTRIBUTES} & \begin{bmatrix} \text{HAS-COLOR} & \text{GREEN} \end{bmatrix} \end{bmatrix}
\end{bmatrix}
$$

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

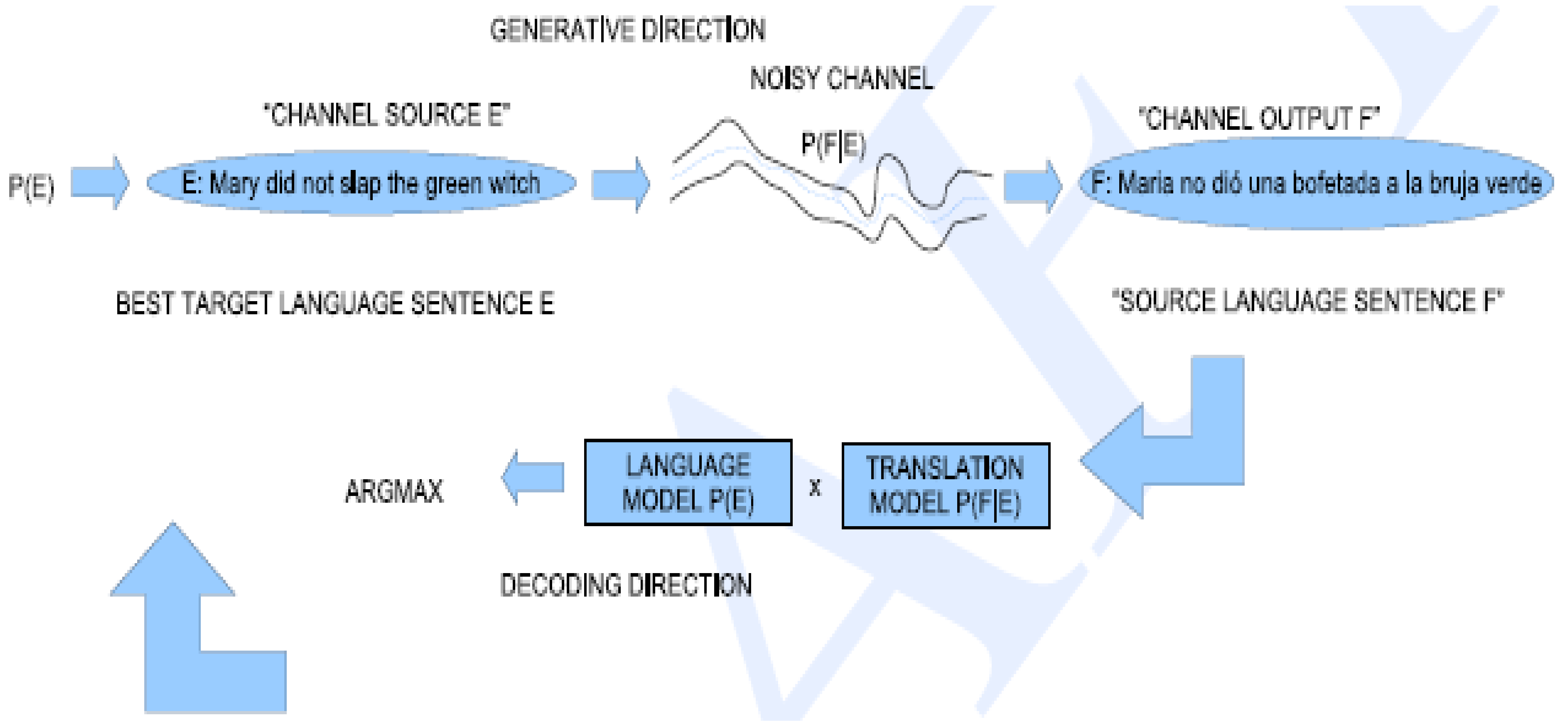
- A good translation should be ***faithful*** and correctly convey the information and tone of the original source sentence.

- A good translation should also be ***fluent***, grammatically well structured and readable in the target language.

- Final objective:

$$T_{best} = \underset{\mathrm{T} \in \mathrm{Target}}{\operatorname{argmax}} \ \mathrm{faithfulness}(T, S) \, \mathrm{fluency}(T)$$

Noisy Channel Model



Bayesian Analysis of Noisy Channel

$$\hat{E} = \underset{E \in English}{\operatorname{argmax}} P(E \mid F)$$

$$= \underset{E \in English}{\operatorname{argmax}} \frac{P(F \mid E) P(E)}{P(F)}$$

$$= \underset{E \in English}{\operatorname{argmax}} \underbrace{P(F \mid E)}_{\text{Translation Model}} \underbrace{P(E)}_{\text{Language Model}}$$

- A decoder determines the most probable translation $\hat{E}$ given F

Language Model

- Use a standard *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 *P* based on simple frequency counts.

$$P(\bar{f},\bar{e}) = \frac{\text{count}(\bar{f},\bar{e})}{\sum_{\bar{f}} \text{count}(\bar{f},\bar{e})}$$

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

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 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.

$$p_n = \frac{\sum_{C \in corpus} \sum_{\text{n-gram} \in C} \text{count}_{\text{clip}}(\text{n-gram})}{\sum_{C \in corpus} \sum_{\text{n-gram} \in C} \text{count}(\text{n-gram})} \qquad p = \sqrt[N]{\prod_{n=1}^{N} p_n}$$

$$\text{Cand 1:} \quad p = \sqrt[2]{\frac{5}{6}\frac{1}{5}} = 0.408$$

$$\text{Cand 2:} \quad p = \sqrt[2]{\frac{7}{10}\frac{4}{9}} = 0.558$$

Brevity Penalty

- Not easy to compute recall since there are multiple references.

- Instead, use a penalty for translations that are shorter than the closest reference.

- 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.

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

BLEU Score

- Final BLEU Score: BLEU = $BP \times p$

Cand 1: Mary no slap the witch green.

Best Ref: Mary did not slap the green witch.

$$c = 6, \quad r = 7, \qquad BP = e^{(1-7/6)} = 0.846$$

$$BLEU = 0.846 \times 0.408 = 0.345$$

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

Best Ref: Mary did not smack the green witch.

$$c = 10, \quad r = 7, \qquad BP = 1$$

$$BLEU = 1 \times 0.558 = 0.558$$