# CS 6120/CS4120: Natural Language Processing

Instructor: Prof. Lu Wang

College of Computer and Information Science

Northeastern University

Webpage: www.ccs.neu.edu/home/luwang

# Machine Translation

- Automatically translate one natural language into another.
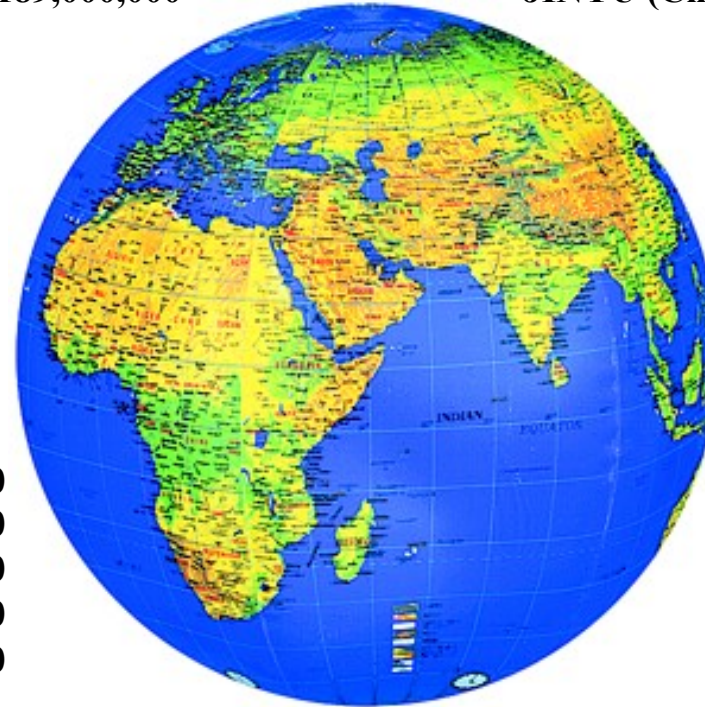
**Mary didn't slap the green witch.**

↓

**Maria no dió una bofetada a la bruja verde.**
**(Mary do not gave a slap to the witch green.)**

[Some slides are borrowed from Raymond Mooney, Kevin Knight, and Alan Ritter]

# Thousands of Languages Are Spoken

| | | | |
|---|---|---|---|
| MANDARIN | 885,000,000 | TURKISH | 59,000,000 |
| SPANISH | 332,000,000 | URDU | 58,000,000 |
| ENGLISH | 322,000,000 | MIN NAN (China) | 49,000,000 |
| BENGALI | 189,000,000 | JINYU (China) | 45,000,000 |
| HINDI | 182,000,000 | GUJARATI | 44,000,000 |
| PORTUGUESE | 170,000,000 | POLISH | 44,000,000 |
| RUSSIAN | 170,000,000 | ARABIC | 42,500,000 |
| JAPANESE | 125,000,000 | UKRAINIAN | 41,000,000 |
| GERMAN | 98,000,000 | | |
| WU (China) | 77,175,000 | | |
| JAVANESE | 75,500,800 | ITALIAN | 37,000,000 |
| KOREAN | 75,000,000 | XIANG (China) | 36,015,000 |
| FRENCH | 72,000,000 | MALAYALAM | 34,022,000 |
| VIETNAMESE | 67,662,000 | HAKKA (China) | 34,000,000 |
| TELUGU | 66,350,000 | KANNADA | 33,663,000 |
| YUE (China) | 66,000,000 | ORIYA | 31,000,000 |
| MARATHI | 64,783,000 | PANJABI | 30,000,000 |
| TAMIL | 63,075,000 | SUNDA | 27,000,000 |

Source: Ethnologue

# 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 do not gave a slap to the witch green.)**
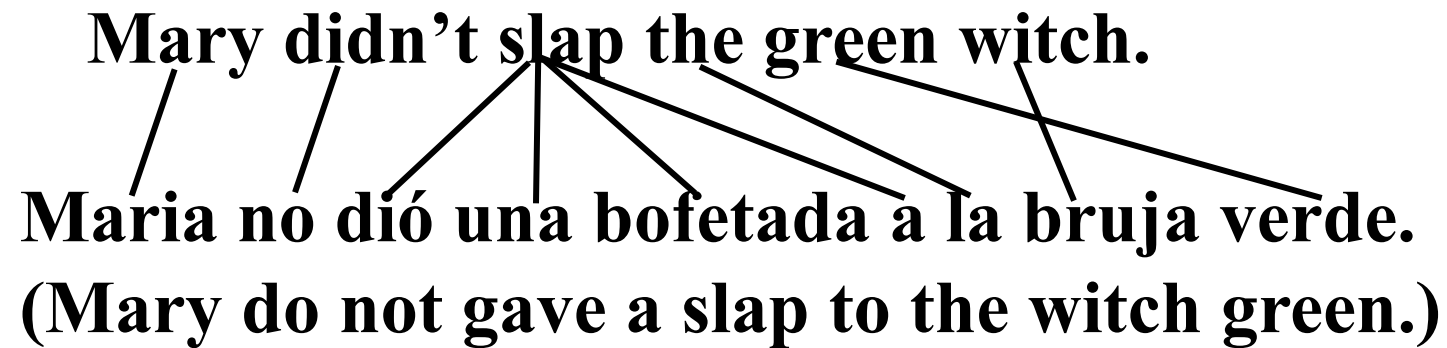
# Translation Quality: what's the current status?

- 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.

# Outline

- Issues in machine translation (MT)

- Direct transfer and syntactic transfer

- Statistical MT and noisy channel model

- MT evaluation

# Ambiguity Resolution is Required for Translation

- Syntactic and semantic ambiguities must be properly resolved for correct translation:
  - "John plays the guitar." → "John toca la guitarra."
  - "John plays soccer." → "John juega el fútbol."
- An apocryphal story is that 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."

# Issues: 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) in Japanese

# Issues: Differing Word Orders

- English word order is subject – verb – object (SVO)
- Japanese word order is subject – object – verb (SOV)

| English: | IBM bought Lotus |
|----------|------------------|
| Japanese: | *IBM Lotus bought* |

| English: | Sources said that IBM bought Lotus yesterday |
|----------|-----------------------------------------------|
| Japanese: | *Sources yesterday IBM Lotus bought that said* |

# Issues: Differing Word Orders

- English word order is subject – verb – object (SVO)
- Japanese word order is subject – object – verb (SOV)

| Word order | English equivalent | Proportion of languages | | Example languages |
|---|---|---|---|---|
| SOV | "She him loves." | 45% | | Sanskrit, Hindi, Ancient Greek, Latin, Japanese, Korean |
| SVO | "She loves him." | 42% | | Chinese, English, French, Hausa, Italian, Malay, Russian, Spanish |
| VSO | "Loves she him." | 9% | | Biblical Hebrew, Arabic, Irish, Filipino, Tuareg-Berber, Welsh |
| VOS | "Loves him she." | 3% | | Malagasy, Baure |
| OVS | "Him loves she." | 1% | | Apalaí, Hixkaryana |
| OSV | "Him she loves." | 0% | | Warao, (certain dialects of) Korean |

**S**ubject, **O**bject, **V**erb

# Issues: Syntactic Structure is not Preserved Across Translations

The bottle floated into the cave

⇓

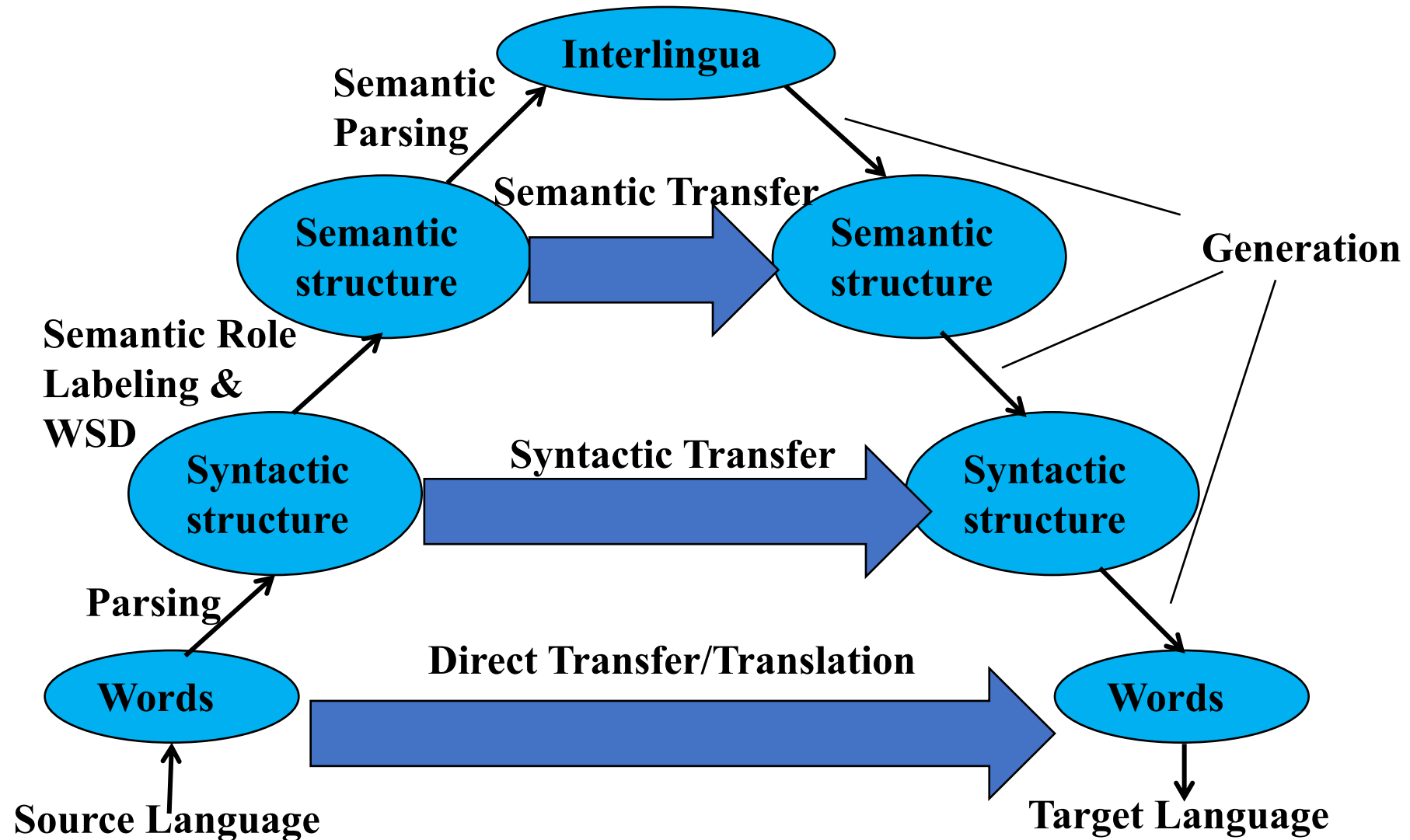La botella entro a la cuerva flotando
(the bottle entered the cave floating)

# Outline

- Issues in machine translation (MT)

- Direct transfer and syntactic transfer

- Statistical MT and noisy channel model

- MT evaluation

# Vauquois Triangle

# Direct Transfer/Translation

- Translation is word-by-word

- Very little analysis of the source text (e.g., no syntactic or semantic analysis)

- Relies on a large bilingual dictionary. For each word in the source language, the dictionary specifies a set of rules for translating that word.

# CLASSIC SOUPS

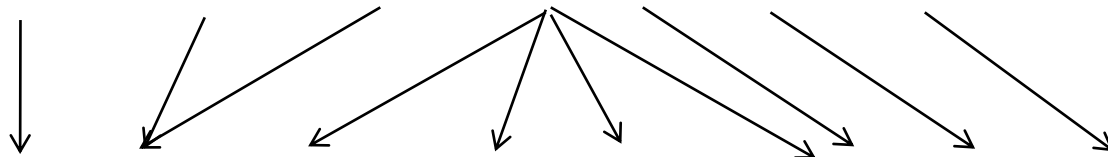| | | Sm. | Lg. |
|---|---|---|---|
| 清燉雞湯 57. | House Chicken Soup (Chicken, Celery, Potato, Onion, Carrot) | 1.50 | 2.75 |
| 雞飯湯 58. | Chicken Rice Soup | 1.85 | 3.25 |
| 雞麵湯 59. | Chicken Noodle Soup | 1.85 | 3.25 |
| 廣東雲吞 60. | Cantonese Wonton Soup | 1.50 | 2.75 |
| 蕃茄蛋湯 61. | Tomato Clear Egg Drop Soup | 1.65 | 2.95 |
| 雲吞湯 62. | Regular Wonton Soup | 1.10 | 2.10 |
| 酸辣湯 63. 🌶 | Hot & Sour Soup | 1.10 | 2.10 |
| 蛋花湯 64. | Egg Drop Soup | 1.10 | 2.10 |
| 雲蛋湯 65. | Egg Drop Wonton Mix | 1.10 | 2.10 |
| 豆腐菜湯 66. | Tofu Vegetable Soup | NA | 3.50 |
| 雞玉米湯 67. | Chicken Corn Cream Soup | NA | 3.50 |
| 蟹肉玉米湯 68. | Crab Meat Corn Cream Soup | NA | 3.50 |
| 海鮮湯 69. | Seafood Soup | NA | 3.50 |

# Direct Transfer/Translation

- 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 no dar:PAST 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.

# An Example of a set of Direct Translation Rules

Rules for translating *much* or *many* into Russian:

**if** preceding word is *how* **return** *skol'ko*
**else if** preceding word is *as* **return** *stol'ko zhe*
**else if** word is *much*
    **if** preceding word is *very* **return** nil
    **else if** following word is a noun **return** *mnogo*
**else** (word is many)
    **if** preceding word is a preposition and following word is noun **return** *mnogii*
    **else return** *mnogo*

# Lack of any analysis of the source language causes several problems

- Difficult or impossible to capture long-range reorderings

English:        Sources said that IBM bought Lotus yesterday

Japanese:      *Sources yesterday IBM Lotus bought that said*

- Words are translated without disambiguation of their syntactic role e.g., that can be a complementizer or determiner, and will often be translated differently for these two cases

They said *that* …

They like *that* ice-cream

# Possible Solution

- Analysis: Analyze the source language sentence; for example, build a syntactic analysis of the source language sentence.

- Transfer: Convert the source-language parse tree to a target-language parse tree.

- Generation: Convert the target-language parse tree to an output sentence.

# 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.

# 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 Nom $\Rightarrow$ NP → Nom ADJ
  - English to Japanese:
    - VP → V NP $\Rightarrow$ VP → NP V
    - PP → P NP $\Rightarrow$ PP → NP P
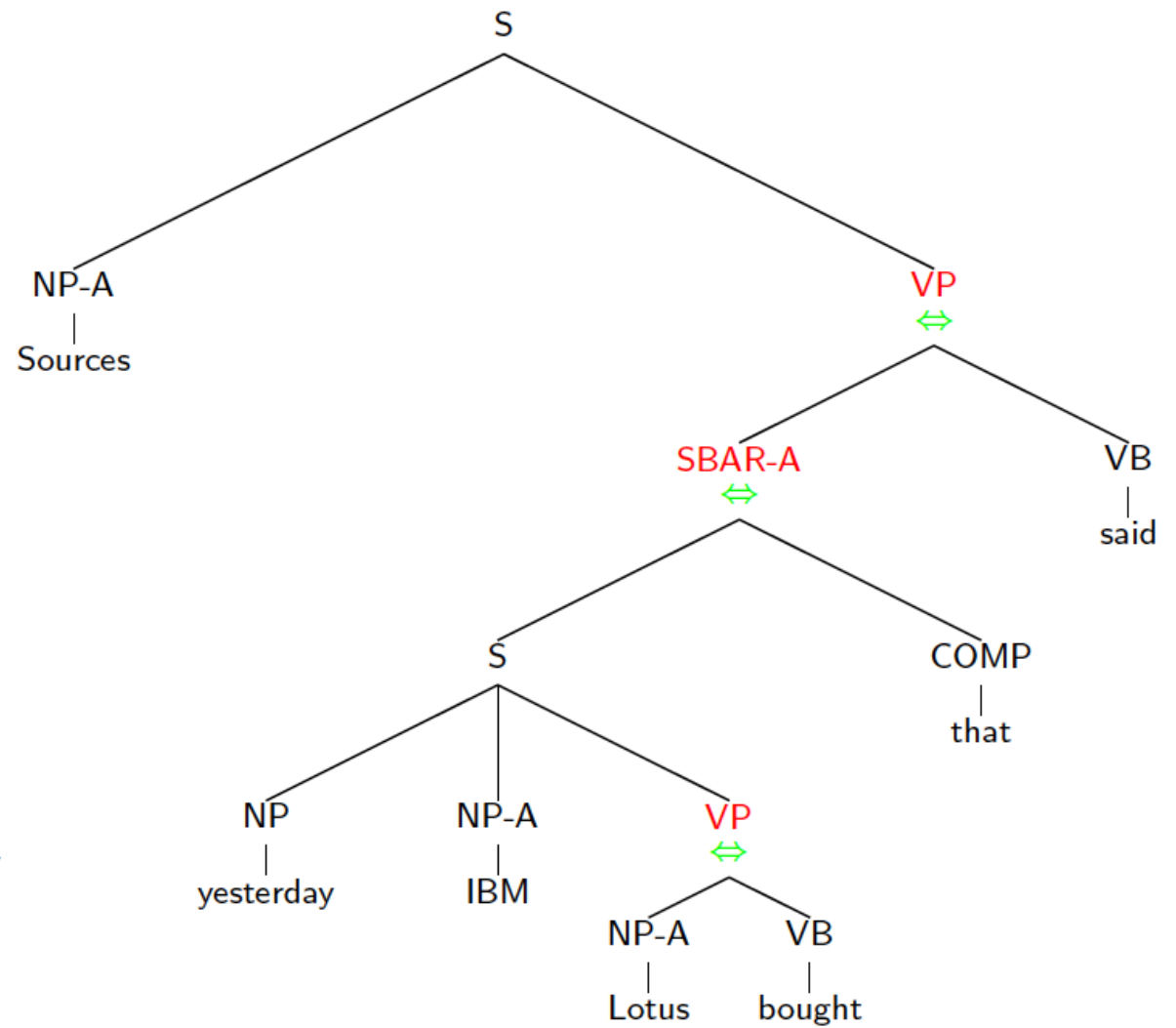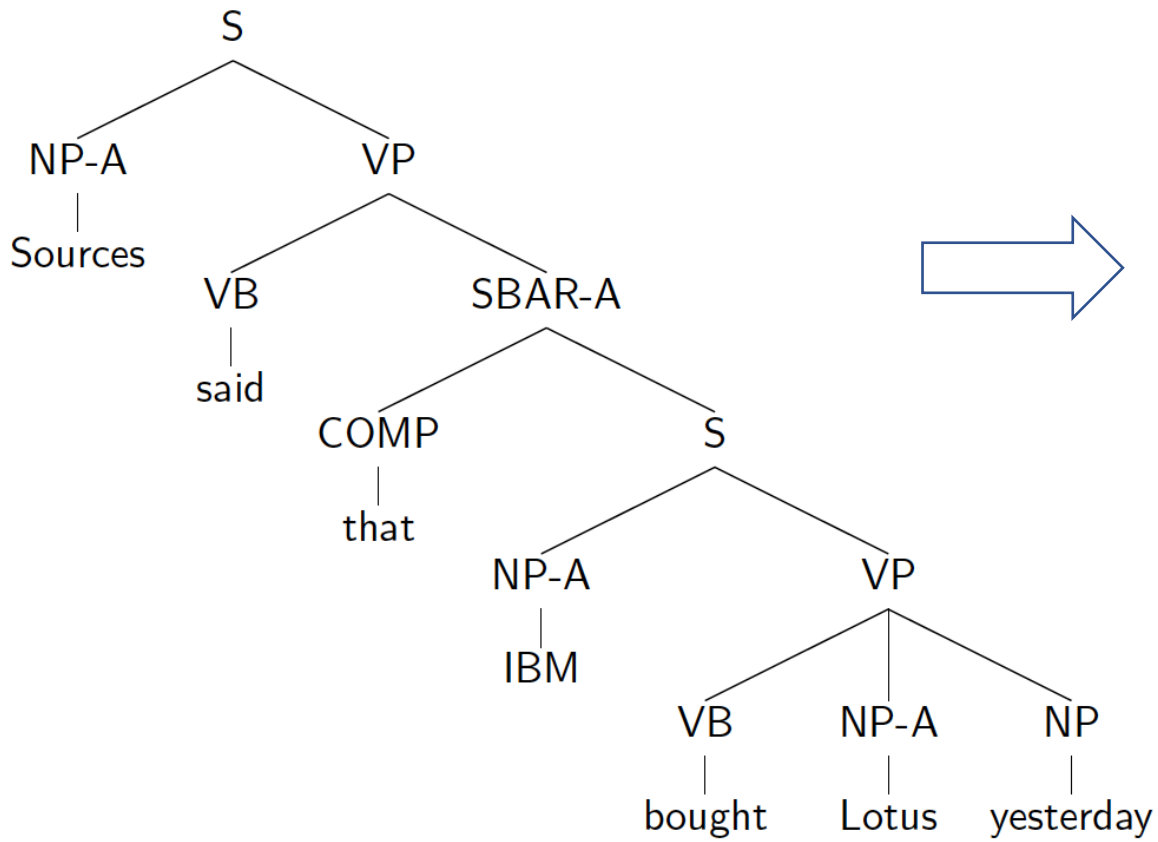
⇒ Japanese: *Sources yesterday IBM Lotus bought that said*

# Outline

- Issues in machine translation (MT)

- Direct transfer and syntactic transfer

→ - Statistical MT and noisy channel model

- MT evaluation

# Statistical MT

- Manually encoding comprehensive bilingual lexicons and transfer rules is difficult.

- SMT acquires knowledge needed for translation from a ***parallel corpus*** or ***bitext*** that contains the same set of documents in two languages.

- The Canadian Hansards (parliamentary proceedings in French and English) is a well-known parallel corpus.

- First align the sentences in the corpus based on simple methods that use coarse cues like sentence length to give bilingual sentence pairs.

- Then align the words in parallel sentences

# Word Alignment



**Mary didn't slap the green witch.**

**Maria no dió una bofetada a la bruja verde.**
**(Mary do not gave a slap to the witch green.)**
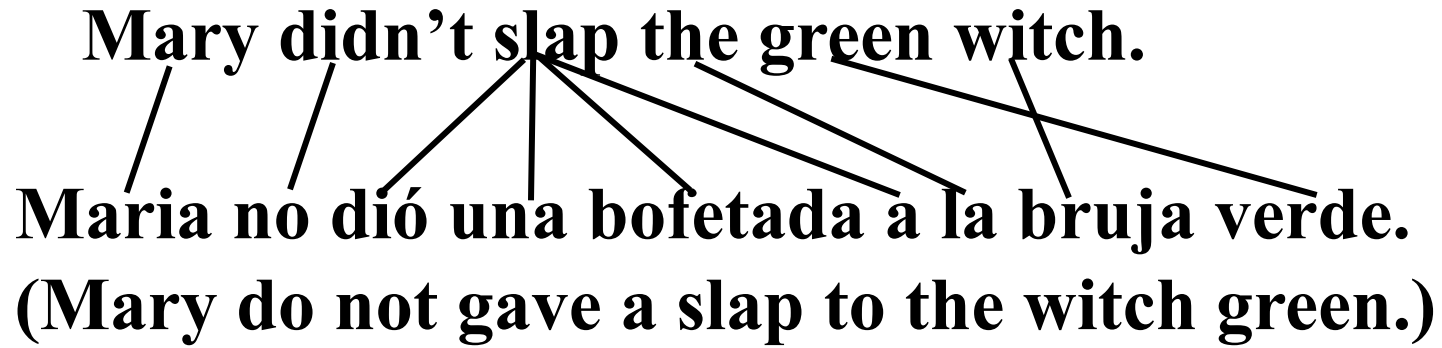
# Word Alignment

- Basic idea: co-occurrence between words and phrases

... la maison ... la maison bleue ... la fleur ...

... the house ... the blue house ... the flower ...
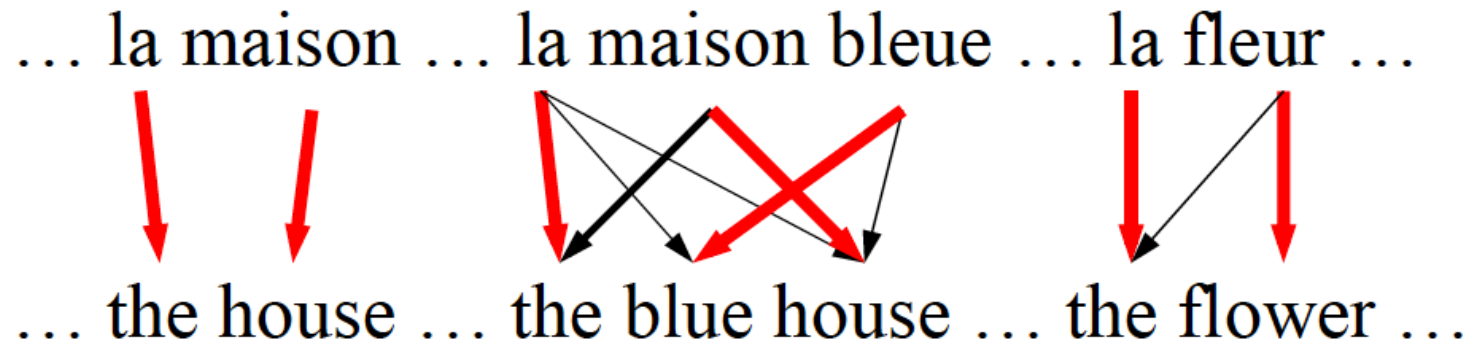
- The IBM models (will not discussed in class, but reference here: http://www.cs.columbia.edu/~mcollins/courses/nlp2011/notes/ibm12.pdf)

| English | French | P(f \| e) |
|---|---|---|
| national | nationale | 0.47 |
|  | national | 0.42 |
|  | nationaux | 0.05 |
|  | nationales | 0.03 |
| the | le | 0.50 |
|  | la | 0.21 |
|  | les | 0.16 |
|  | l' | 0.09 |
|  | ce | 0.02 |
|  | cette | 0.01 |
| farmers | agriculteurs | 0.44 |
|  | les | 0.42 |
|  | cultivateurs | 0.05 |
|  | producteurs | 0.02 |

[Brown et al 93]

# 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:

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

# Noisy Channel Model

- Based on analogy to information-theoretic model used to decode messages transmitted via a communication channel that adds errors.

- Assume that source sentence was generated by a "noisy" transformation of some target language sentence and then use Bayesian analysis to recover the most likely target sentence that generated it.

Translate foreign language sentence $F=f_1, f_2, \ldots f_m$ to an English sentence $\hat{E} = e_1, e_2, \ldots e_I$ that maximizes $P(E \mid F)$

# Bayesian Analysis of Noisy Channel

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

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

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

**Translation Model**    **Language Model**

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

Translation from Spanish to English, candidate translations based on $p(Spanish \mid English)$ alone:

Que hambre tengo yo

$\rightarrow$

| | |
|---|---|
| What hunger have | $p(s\vert e) = 0.000014$ |
| Hungry I am so | $p(s\vert e) = 0.000001$ |
| I am so hungry | $p(s\vert e) = 0.0000015$ |
| Have i that hunger | $p(s\vert e) = 0.000020$ |

. . .

With $p(Spanish \mid English) \times p(English)$:

Que hambre tengo yo

$\rightarrow$

What hunger have $\qquad p(s|e)p(e) = 0.000014 \times 0.000001$

Hungry I am so $\qquad p(s|e)p(e) = 0.000001 \times 0.0000014$

I am so hungry $\qquad p(s|e)p(e) = 0.0000015 \times 0.0001$

Have i that hunger $\quad p(s|e)p(e) = 0.000020 \times 0.0000098$

# Outline

- Issues in machine translation (MT)

- Direct transfer and syntactic transfer

- Statistical MT and noisy channel model

- MT evaluation

# 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 judgements.

# Human Evaluation of MT

- Ask humans to estimate MT output on several dimensions.
  - **Fluency**: Is the result grammatical, understandable, and readable in the target language.
  - **Fidelity**: Does the result correctly convey the information in the original source language.

# 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

# 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: <mark>Mary</mark> <mark>did</mark> <mark>not</mark> give a smack to <mark>a</mark> <mark>green</mark> <mark>witch</mark>.

Ref 1: <mark>Mary</mark> <mark>did not</mark> slap the <mark>green witch</mark>.

Ref 2: <mark>Mary</mark> <mark>did not</mark> smack the <mark>green witch</mark>.

Ref 3: <mark>Mary</mark> <mark>did not</mark> hit <mark>a green</mark> 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, 2 in this example) using geometric mean.

$$p_n = \frac{\displaystyle\sum_{C \in corpus} \sum_{n-gram \in C} count_{clip}(n-gram)}{\displaystyle\sum_{C \in corpus} \sum_{n-gram \in C} count\ (n-gram)}$$

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

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

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

# 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.

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

# BLEU Score

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

<span style="color:red">Cand 1</span>: Mary no slap the witch green.

<span style="color:red">Best Ref</span>: Mary did not slap the green witch.

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

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

<span style="color:red">Cand 2</span>: Mary did not give a smack to a green witch.

<span style="color:red">Best Ref</span>: Mary did not smack the green witch.

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

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

# BLEU Score Issues

- BLEU has been shown to correlate with human evaluation when comparing outputs from different SMT systems.

- However, it is does not correlate with human judgments when comparing SMT systems with manually developed MT (Systran) or MT with human translations.

- Other MT evaluation metrics have been proposed that claim to overcome some of the limitations of BLEU.