Based on slides from Jason Eisenstein, Chris Dyer, Alan Ritter, Yejin Choi, and everyone else they copied from.
Upcoming...

- **Summaries**
  - Paper summaries: **February 28, March 14**
  - Summary 1 graded

- **Project**
  - Status report due in 1 weeks: **March 7, 2017**
  - Instructions coming today!
  - Almost final report, only **5 pages**

- **Homework**
  - Homework 4 is due on **March 13**
  - Write-up and data releasing soon.
Outline

Machine Translation

Introduction to Statistical MT

IBM Translation Models
Outline

Machine Translation

Introduction to Statistical MT

IBM Translation Models
Machine Translation

Yo, que me figuraba el Paraíso / Bajo la especie de una biblioteca.

I, who figured the Paradise / Under the species of a library.

I have always imagined Paradise as a kind of library.
Challenges: Word Order

**SVO vs SOV**

\[
\begin{array}{c}
\text{S} \quad \text{V} \quad \text{O} \\
\end{array}
\]

**English:** IBM bought Lotus
**Japanese:** IBM Lotus bought

**Even for SVO**

**English:** I will buy it
**French:** Je vais l’acheter (I will it buy)

**English:** I bought it
**French:** Je l’ai achet´e (I it have bought)
Challenges: Lexical Ambiguity

en

pico (bird)
es
cuenta (cost)
bill

en

fr

etape

paw

patte

foot

leg

jambe

pied

ANIMAL

HUMAN

CHAIR

BIRD

JOURNEY

ANIMAL
Challenges: Pronouns

English possessive pronouns take the gender of the owner:
Marie rides her bike

French possessive pronouns take the gender of the object:
Marie monte sur son vélo

In Spanish, you can recover the pronoun from verb inflection:
Vivimos en Atlanta → We live in Atlanta

Again, discourse context is often crucial:
Vive en Atlanta → She/he/it lives in Atlanta

这块蛋糕很美味。谁烤的？
Zhè kuài dāngāo hěn méiwèi. Shéi kāo de?
This piece cake very beautiful taste. Who bake?
"This cake is very tasty. Who baked it?"
Challenges: Tenses

The preterite tense is for events with a definite time, e.g.

I biked to work this morning

The imperfect is for events with indefinite times, e.g.

I biked to work all last summer

To translate English to Spanish, we must pick the right tense.
Challenges: Idioms

- Why in the world
- Hold Your Horses
- Kick the bucket
- Bob's Your Uncle
- Storm in a Teacup
- As Cool As a Cucumber
- Lend me your ears
- Dead As A Doornail
- Blue in the Face
- Head In The Clouds
Rules for Machine Translation

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

Panov (1960)
The Vauquios Triangle

"Vauquois Triangle"

| English (E) | P( E | lo haré ) |
|-------------|-------------|
| will do it  | 0.8         |
| will do so  | 0.2         |

| English (E) | P( E | mañana ) |
|-------------|-------------|
| tomorrow    | 0.7         |
| morning     | 0.3         |
Outline

- Machine Translation
- Introduction to Statistical MT
- IBM Translation Models
Statistical Machine Translation

Sentence-aligned parallel corpus:

- Yo lo haré mañana
  - I will do it tomorrow
- Hasta pronto
  - See you soon
- Hasta pronto
  - See you around

Machine translation system:

- Yo lo haré pronto
  - Novel Sentence
- Model of translation

- I will do it soon
- I will do it around
- See you tomorrow
Parallel Corpus: Examples
Parallel Corpus: Examples
Parallel Corpus: Examples
Parallel Corpus: Examples

<table>
<thead>
<tr>
<th>No.</th>
<th>Description</th>
<th>Sm.</th>
<th>Lg.</th>
</tr>
</thead>
<tbody>
<tr>
<td>57.</td>
<td>House Chicken Soup (Chicken, Celery, Potato, Onion, Carrot)</td>
<td>1.50</td>
<td>2.75</td>
</tr>
<tr>
<td>58.</td>
<td>Chicken Rice Soup</td>
<td>1.85</td>
<td>3.25</td>
</tr>
<tr>
<td>59.</td>
<td>Chicken Noodle Soup</td>
<td>1.85</td>
<td>3.25</td>
</tr>
<tr>
<td>60.</td>
<td>Cantonese Wonton Soup</td>
<td>1.50</td>
<td>2.75</td>
</tr>
<tr>
<td>61.</td>
<td>Tomato Clear Egg Drop Soup</td>
<td>1.65</td>
<td>2.95</td>
</tr>
<tr>
<td>62.</td>
<td>Regular Wonton Soup</td>
<td>1.10</td>
<td>2.10</td>
</tr>
<tr>
<td>63.</td>
<td>Hot &amp; Sour Soup</td>
<td>1.10</td>
<td>2.10</td>
</tr>
<tr>
<td>64.</td>
<td>Egg Drop Soup</td>
<td>1.10</td>
<td>2.10</td>
</tr>
<tr>
<td>65.</td>
<td>Egg Drop Wonton Mix</td>
<td>1.10</td>
<td>2.10</td>
</tr>
<tr>
<td>66.</td>
<td>Tofu Vegetable Soup</td>
<td>NA</td>
<td>3.50</td>
</tr>
<tr>
<td>67.</td>
<td>Chicken Corn Cream Soup</td>
<td>NA</td>
<td>3.50</td>
</tr>
<tr>
<td>68.</td>
<td>Crab Meat Corn Cream Soup</td>
<td>NA</td>
<td>3.50</td>
</tr>
<tr>
<td>69.</td>
<td>Seafood Soup</td>
<td>NA</td>
<td>3.50</td>
</tr>
</tbody>
</table>
The Rosetta Stone

 Hieroglyphics

 Demotic

 Egyptian

 Greek
Warren Weaver (1949)

One naturally wonders if the problem of translation could conceivably be treated as a problem in cryptography. When I look at an article in Russian, I say: ‘This is really written in English, but it has been coded in some strange symbols. I will now proceed to decode.’
Parallel Corpus: Examples

Harry Potter and the Sorcerer's Stone

哈利·波特与魔法石
## Parallel Corpus: Examples

### Classic Soups

<table>
<thead>
<tr>
<th>No.</th>
<th>Name</th>
<th>Sm.</th>
<th>Lg.</th>
</tr>
</thead>
<tbody>
<tr>
<td>57.</td>
<td>House Chicken Soup (Chicken, Celery, Potato, Onion, Carrot)</td>
<td>1.50</td>
<td>2.75</td>
</tr>
<tr>
<td>58.</td>
<td>Chicken Rice Soup</td>
<td>1.85</td>
<td>3.25</td>
</tr>
<tr>
<td>59.</td>
<td>Chicken Noodle Soup</td>
<td>1.85</td>
<td>3.25</td>
</tr>
<tr>
<td>60.</td>
<td>Cantonese Wonton Soup</td>
<td>1.50</td>
<td>2.75</td>
</tr>
<tr>
<td>61.</td>
<td>Tomato Clear Egg Drop Soup</td>
<td>1.65</td>
<td>2.95</td>
</tr>
<tr>
<td>62.</td>
<td>Regular Wonton Soup</td>
<td>1.10</td>
<td>2.10</td>
</tr>
<tr>
<td>63.</td>
<td>Hot &amp; Sour Soup</td>
<td>1.10</td>
<td>2.10</td>
</tr>
<tr>
<td>64.</td>
<td>Egg Drop Soup</td>
<td>1.10</td>
<td>2.10</td>
</tr>
<tr>
<td>65.</td>
<td>Egg Drop Wonton Mix</td>
<td>1.10</td>
<td>2.10</td>
</tr>
<tr>
<td>66.</td>
<td>Tofu Vegetable Soup</td>
<td>NA</td>
<td>3.50</td>
</tr>
<tr>
<td>67.</td>
<td>Chicken Corn Cream Soup</td>
<td>NA</td>
<td>3.50</td>
</tr>
<tr>
<td>68.</td>
<td>Crab Meat Corn Cream Soup</td>
<td>NA</td>
<td>3.50</td>
</tr>
<tr>
<td>69.</td>
<td>Seafood Soup</td>
<td>NA</td>
<td>3.50</td>
</tr>
</tbody>
</table>
Noisy Channel Model

\[
y' = \arg\max_y p(y|x) = \arg\max_y \frac{p(x|y)p(y)}{p(x)} = \arg\max_y p(x|y)p(y)
\]
Noisy Channel Model

\[ e \overset{\text{en}}{\rightarrow} e' \]

\[ e' = \arg\max_e p(f|e) p(e) \]

Translation Model

Language Model
Example: Noisy Channel

Que hambre tengo yo

→

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

...
Example: Noisy Channel

Que hambre tengo yo

→

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

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

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

Have i that hunger \( p(s|e)p(e) = 0.000020 \times 0.00000098 \)

...
Components of an MT system

Language Model

Translation Model

Decoding Algo

\[ p(e) \]

\[ p(\text{fle}) \]

\[ \arg\max_e p(\text{fle}) \cdot p(e) \]
Components of an MT system

S'il vous plait traduire...

decoder

\[ e' = \operatorname{arg\,max}_e p(e \mid f) \]

Please translate...

Learner

English

Learner

TM

\[ p(f \mid e) \]

Learner

English

français
Evaluating MT

More has been written about machine translation evaluation than about machine translation itself.

- Yorrick Wilks
Human Evaluation

```
"Does e look like english?"
"Does it mean same as french?"
```

A: furious nAgA on wednesday, the tribal minimum pur of ten schools also was burnt

B: furious nAgA on wednesday the tribal pur mini ten schools of them was also burnt
Automated Evaluation

\[ f \rightarrow e \]

Fluency

Adequacy

\[ f \rightarrow \{ e_1, e_2, e_3 \} \rightarrow e \]
BLEU Score

$0.0 \rightarrow 1.0$

$e \sim 1\text{-grams}$\hspace{1cm} \text{prec}(1)$

$\text{uni}(e) \cap \text{uni}(e,e_2..e_n)$

$\sim 2\text{-grams}$

$\vdots$

$\text{BLEU}_4 = \left\lbrack \prod_{k=1}^{4} \text{prec}(k)^{1/4} \right\rbrack \left\lbrack \begin{array}{c} \text{Brevity} \\ \text{Penalty} \end{array} \right\rbrack$
BLEU Score: Example

\[ p_{\text{prec}}(1) = \frac{1 + 1 + 1 + 1 + 1 + 1 + 1}{1 + 1 + 1 + 1 + 1 + 1 + 1} = \frac{7}{8} = 0.875 \]
BLEU Score: Example

‘extension of isi in uttar pradesh ’
‘ isi ’s expansion in uttar pradesh ’
‘ the spread of isi in uttar pradesh ’
‘ isi spreading in uttar pradesh ’
the spread of isi in uttar pradesh

\[ \text{prec}(2) = \frac{5}{7} \quad \text{prec}(3) = \frac{4}{6} \quad \text{prec}(4) = \frac{3}{5} \]
BLEU’s not bad...

![Graph showing the correlation between NIST Score (variant of BLEU) and Human Judgments with Adequacy and Fluency markers. The graph indicates high correlation with R² values of 0.98 and 0.90.](image-url)
Outline

1. Machine Translation
2. Introduction to Statistical MT
3. IBM Translation Models
Statistical Translation Model

And the program was implemented

La programmation a été mise en application

\[ p(fle) = \frac{p(tal\ the) \cdot p(programme) \cdot p(program)}{p(tal\ the) \cdot p(programme) \cdot p(program)} \]
Word Alignment: Direct

Alignment Function

\[ a = \{1 \rightarrow 1, 2 \rightarrow 2, 3 \rightarrow 3, 4 \rightarrow 4\} \]
Word Alignment: 1-to-Many

\[
\begin{align*}
\text{das} &\rightarrow \text{the} \\
\text{Haus} &\rightarrow \text{house} \\
\text{ist} &\rightarrow \text{is} \\
\text{klitzeklein} &\rightarrow \text{very small}
\end{align*}
\]

\[a = \{1 \rightarrow 1, 2 \rightarrow 2, 3 \rightarrow 3, 4 \rightarrow 4, 5 \rightarrow 4\}\]
Word Alignment: Reordering

\[ a = \{ 1 \rightarrow 3, 2 \rightarrow 4, 3 \rightarrow 2, 4 \rightarrow 1 \} \]
Word Alignment: Inserting
Word Alignment: Dropping

```
1 2 3 4
das Haus ist klein

1 2 3
house is small
```
Translating with Alignments

\[ P(f|e) = \sum_{a} P(f, a|e) \]

\[ = \prod_{i} \sum_{\alpha_i} q(a_i|i, f, m) \sum_{t(f_i|e, \alpha_i)} \]
Example: Translation Prob

\[ P(f, a | e) = q(2 | 1, 6, 7) \times t(\text{La | the}) \]
\[ \times q(3 | 2, 6, 7) \times t(\text{Programmation | program}) \]
\[ \times q(4 | 3, 6, 7) \times t(a | has) \]
\[ \times q(5 | 4, 6, 7) \times t(\text{été | been}) \]
\[ \times q(6 | 5, 6, 7) \times t(\text{mise | implemented}) \]
\[ \times q(6 | 6, 6, 7) \times t(\text{en | implemented}) \]
\[ \times q(6 | 7, 6, 7) \times t(\text{application | implemented}) \]
IBM Models

Model 1

\[ q(j \mid i, l, m) = \frac{1}{k} \]

Model 2

\[ q(j \mid i, l, m) = \frac{c(j, i, l, m)}{c(i, l, m)} \]

Model 3/4/5

HMMs

\[ q(j-1) \approx q(j) \]

add new words...
Word Alignment Algorithm

\[ a^* = \arg\max_a p(a|f,e) \]

is simple

\[ a_i = \arg\max_j q(j | l_i, l, m) t(f_l, e_j) \]