MACHINE TRANSLATION
II: NOW WITH 100%
MORE SYNTAX

CS 562/662: Natural Language Processing
2015-03-05
IBM MODEL III
MODEL III

Mary did not slap the green witch

Mary not slap slap slap the green witch \((P_f)\)

Mary not slap slap slap \(\emptyset\) the green witch \((P_\emptyset)\)

Maria no dió una bofetada a la verde bruja \((P)\)

Maria no dio una bofetada a la bruja verde \((P_d)\)

[h/t: Kevin Knight]
\( P_f(n \mid s) \): target token \( s \) aligns to \( n \) source tokens

\( P(t \mid \emptyset) \): a target token \( t \) aligns to no source token

\( P(t \mid s) \): target token \( t \) is generated by aligned source token \( s \)

\( P_d(j \mid i, |S|, |T|) \): target token \( t \) appears in position \( j \) when it is generated by aligned source token in position \( i \) and the source and target are \(|S|\) and \(|T|\) long, respectively

\( P(t_0\ldots t_{|T|}) \): the target consists of \( t_0\ldots t_{|T|} \)
Mary did not slap the green witch.

Maria no dio una bofetada a la bruja verde.
MODEL III PARAMETER ESTIMATION

• One candidate alignment:

\[ f_d(8 \mid 5, 7, 9) = f_d(8 \mid 5, 7, 9) + 1 \ldots \]

• Two candidate alignments:

\[ f_d(8 \mid 5, 7, 9) = f_d(8 \mid 5, 7, 9) + 1/2 \]
\[ f_d(8 \mid 6, 7, 9) = f_d(8 \mid 6, 7, 9) + 1/2 \]

• But, the set of possible alignments grows very fast
MODEL I: TRANSLATION MODEL
ESTIMATION VIA THE E.M. ALGORITHM

1. Compute $f(t \mid s)$, the conditional frequency of $s$ co-occurring with $t$

2. Normalize $f(t \mid s)$ to get $P(t \mid s)$, the maximum likelihood conditional probability distribution

3. Let $a = 0$, $T = 0$ for all arguments.

4. For $n$ iterations:
   
   1. For all $s$, $t$, let

      \[ a(s, t) = a(s, t) + P(t \mid s) \]
      \[ Z(t) = Z(t) + P(t \mid s) . \]

   2. For all $s$, $t$, let

      \[ P(t \mid s) = a(s, t) / Z(t) \]

      then renormalize $P(t \mid s)$. 

MODEL III: ALIGNMENT MODEL ESTIMATION VIA THE E.M. ALGORITHM

1. For $n$ iterations:

   A. For all $s_i$, $t_j$, let

   $$a(s_i, t_j) = a(s_i, t_j) + P(t_j \mid s_i) P_d(j \mid i, |S|, |T|)$$
   $$Z(t_j) = Z(t_j) + P(t_j \mid s_i) P_d(j \mid i, |S|, |T|).$$

   B. For all $s_i$, $t_j$, let

   $$P(t_j \mid s_i) = a(s_i, t_j) / Z(t_j)$$

   then renormalize $P$ and $P_d$. 
Uniform $t$ values

Model 1 iteration
collect fractional counts over all alignments

Revised $t$ values

Uniform $a$ values

Model 2 iteration
collect fractional counts over all alignments

Revised $t, a$ values

Model 2 → 3
one time only transfer of parameter values

Revised $t, a$ values
Initial $n, d, p$ values

Model 3
collect fractional counts over a subset of alignments, identified with the help of the Model 2 Viterbi alignment

Revised $t, a, n, d, p$ values

Final $t, a, n, d, p$ values
PHRASE-BASED TRANSLATION
BASIC PHRASE-BASED TRANSLATION MODELS

1. Segment source $S$ into phrases $s_1 \ldots s_N$

2. Reorder each $s_i$ according to distortion model $P_d$

3. Translate each $s_i$ according to phrasal translation model $P$

[Och & Ney 2004]
Maria no [dio una bofetada] a la bruja verde
Mary did not slap the green witch
PHRASAL ALIGNMENT
TEMPLATES WITH GAPS

… ne voudrais pas voyager par chemin de fer
… would not like traveling by railroad

[Bansal et al. 2011]
SYNTAX-BASED TRANSLATION
HIERARCHICAL PHRASAL ALIGNMENT

he adores [ listening [ to music ] ]

kare ha [[ ongaku wo ] ] kiku ] no ga daisuki desu

[Knight & Yamada 2001, Chiang 2005]
VAUQUOIS TRIANGLE
DECODING AS PARSING

1. Pick target tokens $T$

2. Project pre-terminals over tokens in $T$

3. Combine pre-terminals to project higher-level non-terminal until all tokens in $T$ share a root
he likes listening to music.