

# The Complexity of Phrase Alignment Problems

John DeNero and Dan Klein  
 {denero, klein}@cs.berkeley.edu



## Problem

### Finding Optimal Phrase Alignments

#### Problem Instance

Assigns a weight to every possible phrase alignment

Japan	to	freeze	aid	to	Russia	.	Gloss
0.9		0.01		0.001			日本 Japan
		0.6					冻结 freeze
			0.002				向 to
							俄 Russia
		0.001		0.1			提供 supply
							援助 assistance
			0.0003				。
						0.9	。

#### Optimal Solution

The highest scoring one-to-one and onto phrase alignment

Japan	to	freeze	aid	to	Russia	.	Gloss
0.9							日本 Japan
		0.6					冻结 freeze
				0.8			向 to
						0.9	俄 Russia
			0.7				提供 supply
							援助 assistance
						0.9	。
							。

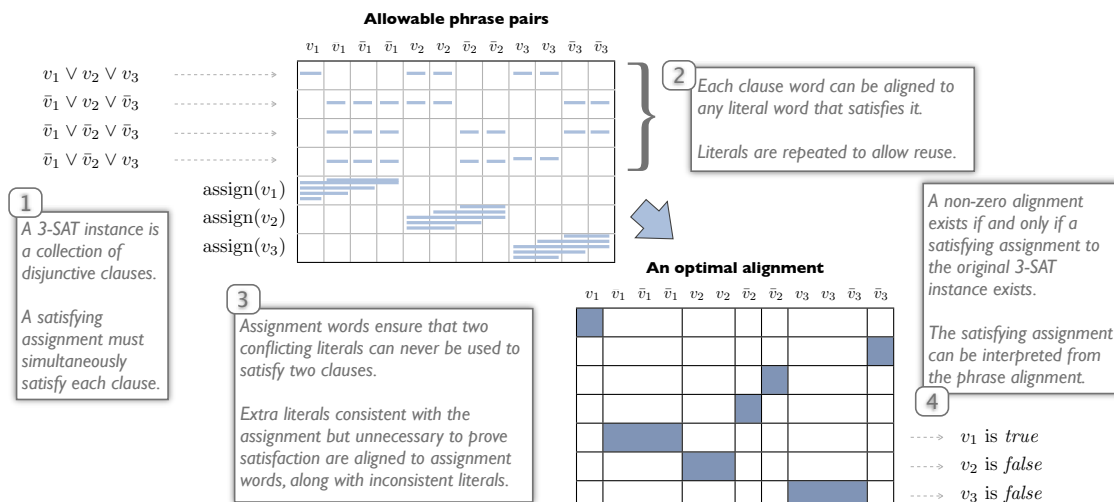
$$0.9 \cdot 0.6 \cdot 0.7 \cdot 0.8 \cdot 0.9 \cdot 0.9 = 0.24$$

### Applications in Machine Translation

- Inference under a phrase alignment model
- Viterbi EM training for a phrase alignment model
- Forced decoding for phrase-based systems

## Analysis & Approximation

### Phrase Alignment is NP-hard by Reduction from Satisfiability



### Phrase Alignment can be Cast as an Integer Linear Program

#### Constants and Variables in the Program

Indicator variables describe sentence segmentation and alignment

Japan	to	freeze	aid	to	Russia	.
0	1	2	3	4	5	6
1						
2						
3						
4						
5						
6						
7						

Constants:  
 $w_{2,5,1,3} = \log 0.02$

Indicator Variables:  
 $a_{2,5,1,3} = 1$   
 $e_{2,5} = 1$   
 $f_{1,3} = 1$

#### Expressing the Objective and Constraints

Constraints enforce segmentation and alignment consistency

Japan	to	freeze	aid	to	Russia	.
0	1	2	3	4	5	6
1						
2						
3						
4						
5						
6						
7						

max  $\sum_{i,j,k,l} w_{i,j,k,l} \cdot a_{i,j,k,l}$

s.t.  $\sum_{i,j:i < x \leq j} e_{i,j} = 1$

$\sum_{k,l:k < y \leq l} f_{k,l} = 1$

$\sum_{k,l} a_{i,j,k,l} = e_{i,j}$

$\sum_{i,j} a_{i,j,k,l} = f_{k,l}$