The Complexity of Phrase Alignment Problems

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

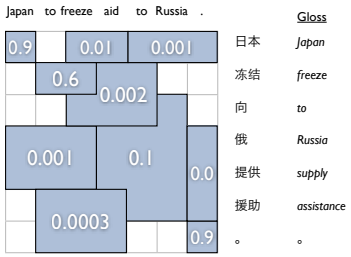
University of California Berkeley

Problem

Finding Optimal Phrase Alignments

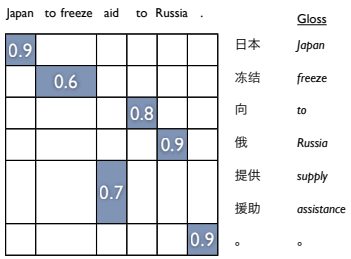
Problem Instance

Assigns a weight to every possible phrase alignment



Optimal Solution

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



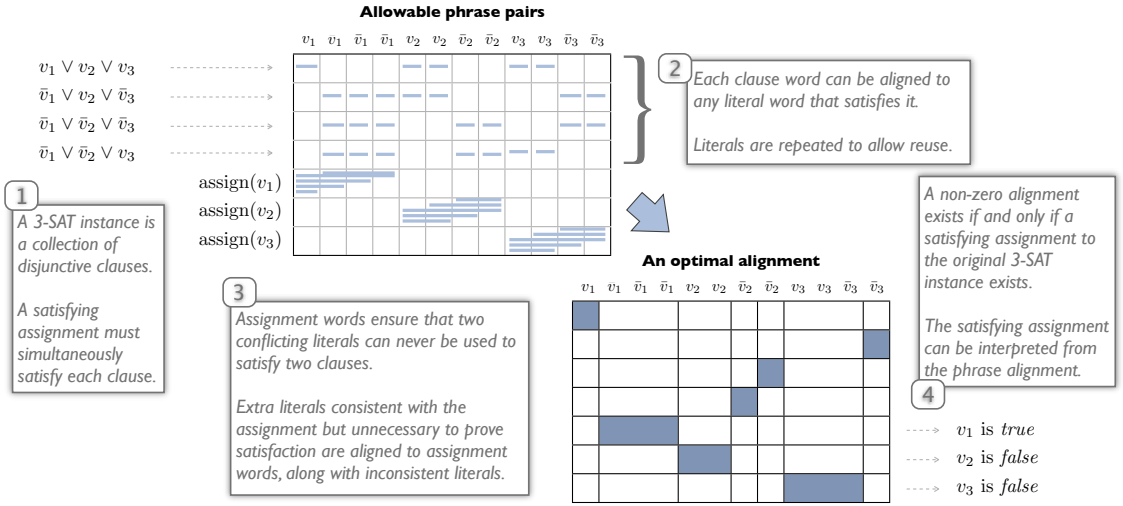
0.9 • 0.6 • 0.7 • 0.8 • 0.9 • 0.9 = 0.24

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

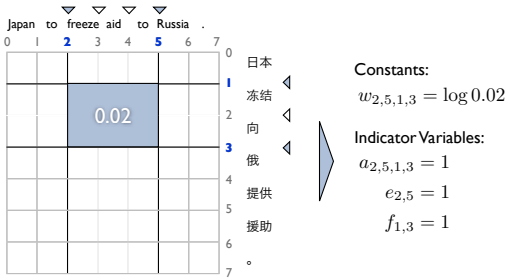
Phrase Alignment is NP-hard by Reduction from Satisfiability



Phrase Alignment can be Cast as an Integer Linear Program

Constants and Variables in the Program

Indicator variables describe sentence segmentation and alignment



Expressing the Objective and Constraints

Constraints enforce segmentation and alignment consistency