北京交通大学
BEIJING JIAOTONG UNIVERSITY

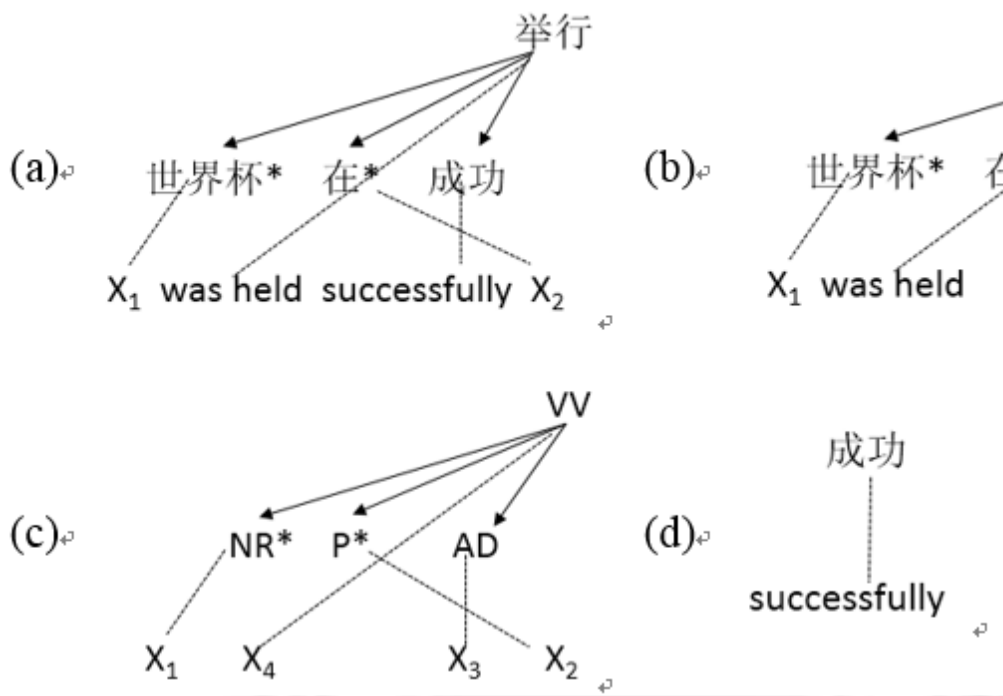
A Dependency-to-String Model for Chinese-Japanese SMT System

Hua Shan Lu Bai Te Luo Yujie Zhang

Beijing jiaotong University

北京交通大学 BEIJING JIAOTONG UNIVERSITY

Dependency-to-String Grammar



- (a), (b) and (c) is HDR rules, (d) is H rules
- HDR rules: the source side is generalized HDR fragments and the target side is strings.
- H rules: the source side is a word and the target side is words or strings.

Rule Acquisition

- Tree annotation
  - Annotate the necessary information on each node of dependency trees for translation rule acquisition
- Identification of acceptable HDR fragments
  - Identify HDR fragments from the annotated trees for HDR rules generation
- HDR rules generation
  - Generate a set of HDR rules according to the identified acceptable HDR fragments

Decoding

- Algorithm
  - Bottom up chart parsing
- Goal
  - Find the best derivation among all possible derivations
- Procedure
  - Apply H rules when n is leaf node
  - Apply HDR rules when n is an internal node
  - Generate the candidate translation for n by cube pruning algorithm

Experiment and Evaluation

- Chinese processing
  - Stanford Word Segmenter
  - Stanford Parser
- Japanese processing
  - JUMAN
  - SRI Language Modeling Toolkit

System	Rule #	BLEU	RIBES
Baseline	35M	34.25	78.94
Ours	8.8M	34.87	79.25

- Baseline: MOSES PBSMT system
- Ours performed better although using only a small size of translation rules

北京交通大學
BEIJING JIAOTONG UNIVERSITY

Thank you!