ISI's 2005 Statistical Machine Translation Entries

Steve DeNeefe and Kevin Knight

Information Sciences Institute
University of Southern California

Outline

- Overview of two MT systems
- Syntax-based Translation Model
- Language Model
- Model Weight Training
- Syntax-based Decoder
- Decoding Example
- Results
- Discussion

Overview

- ISI's two statistical MT systems this year:
  - a phrase-based system
    - intended to be representative of current state-of-the-art techniques in MT
    - poor performance due to user error (OOPS!)
  - a syntax-based system
    - a current research effort at ISI
    - performance is steadily improving

Phrase-based MT system

- nothing new here, really
  - statistical model trained by learning phrase pairs from bilingual data
  - log-linear model allows combination with other knowledge sources (e.g. trigram LM)
  - parameter tuning required for best results
  - rule-based preprocessing for translating dates, numbers, etc.
  - translation model is string-to-string

Phrase-based MT system

- “small” problem during evaluation
  - phrase tables not collected correctly with respect to the evaluation source text
  - thus, our system did not have all the relevant phrase-pairs while decoding

Syntax-based MT system

- similarities to phrase-based system
  - statistical model trained by learning "translation rules" from bilingual data
  - log-linear model allows combination with other knowledge sources (e.g. trigram LM)
  - parameter tuning required for best results
  - rule-based preprocessing for translating dates, numbers, etc.

Syntax-based MT system

- differences from phrase-based system
  - translation model incorporates syntactic structure on the target language side
  - the decoder uses a parser-like method to create syntactic trees as output hypotheses
  - tree-to-string translation model

Syntax-based Translation Model

- rules translate source language phrase into target language syntactic chunks:
  - NPB(PRP/I) ↔ 我
  - NN/hotel ↔ 酒店
  - NP-C(NPB(DT/this NN/address)) ↔ 这个 地址

Syntax-based Translation Model

- rules can have "holes" in the phrases:
  - NP-C(NPB(PRP$/my x_0:NN)) ↔ 我 的 x_0
  - NP-C(NPB(PRP$/my x_0:NN)) ↔ 我 x_0
  - PP(TO/to NP-C(NPB(x_0:NNS NNP/park))) ↔ 去 x_0 公园

Syntax-based Translation Model

- rules can combine previously translated results together:
  - VP(x_0:VBZ x_1:NP-C) $\leftrightarrow$ x_0 x_1
    - combines a verb and a noun-phrase to build a new verb phrase
  - VP(x_0:VBZ x_1:NP-C) $\leftrightarrow$ x_1 x_0
    - takes a noun phrase followed by a verb, switches their order, then combines them into a new verb phrase

Learning the rules

- four steps:
  1. word-align a bilingual parallel corpus
     - union of GIZA++ alignments in each direction
  2. parse the target side
     - using our own implementation of Collins Model 2
  3. extract a list of translation rules
     - using GHKM algorithm (Galley et al, 2004)
  4. estimate probabilities according to relative frequency
     - rule probabilities are conditioned only on root of target syntax fragment – basically a joint p(e,f) model

Language Model

- all language models created with SRI toolkit on English portion of supplied data
- evaluation run
  - bigram model integrated into decoder search
  - 25,000 n-best list re-ranked with trigram model
- post-eval run
  - trigram model integrated into decoder search

Model Weight Training

- split provided development data into dev and test sets:
  - Chinese, Arabic, and Japanese:
    - devset 1 (CSTAR 03) for testing
    - devset 2 (IWLST 04) for development
  - Korean
    - first half of devset 1 (CSTAR 03) for testing, second half for development

Model Weight Training

- parameters trained for syntax system
  - translation model – p(e, f)
  - IBM model 1 inverse approximation
  - language model
  - length bonus and rule bonus
- used exhaustive method to train weights
  - run the decoder on the development set using hundreds of parameter settings, measure BLEU score for each, then use the best one
  - this is time intensive – we only did this for Chinese, and used the results for other languages

Syntax-based Decoder

- probabilistic CYK-style parsing algorithm with beams
- results in an English syntax tree corresponding to the Chinese sentence
- guarantees the output to have some kind of globally coherent syntactic structure

Decoding Example

我 不 懂 英语 .

Literally: “I not understand English .”

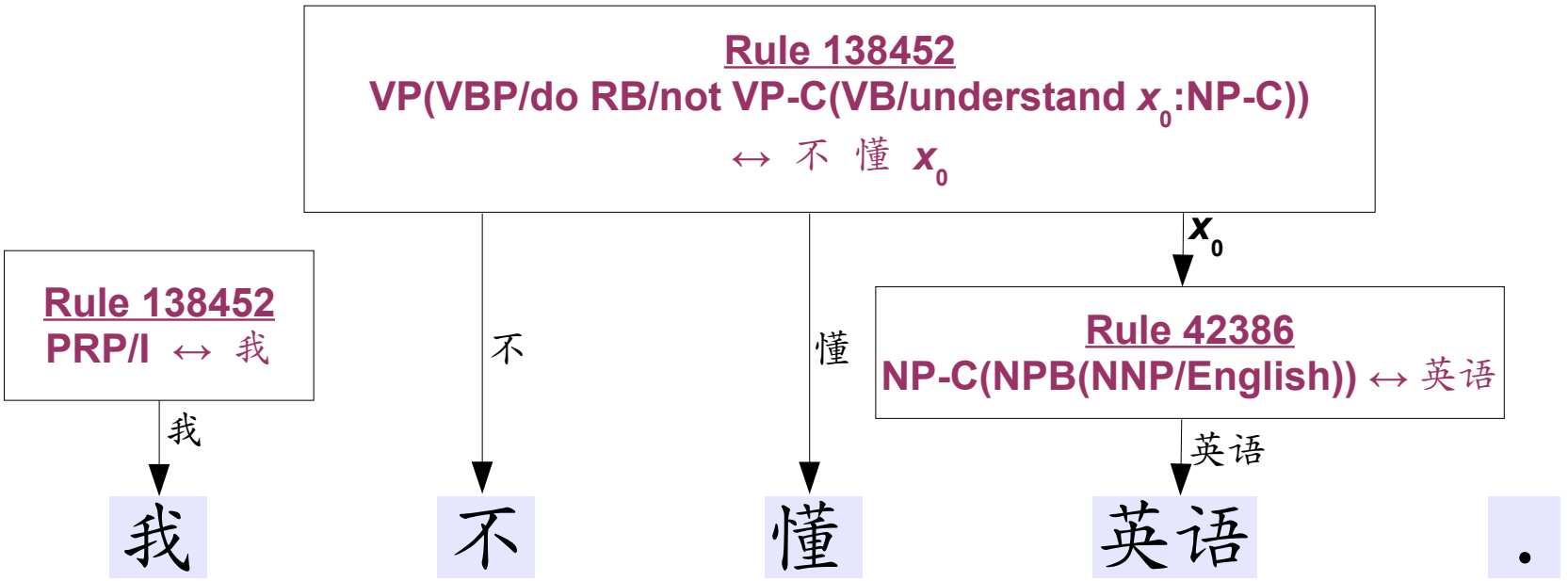
Decoding Example

Literally: “I not understand English .”

Decoding Example

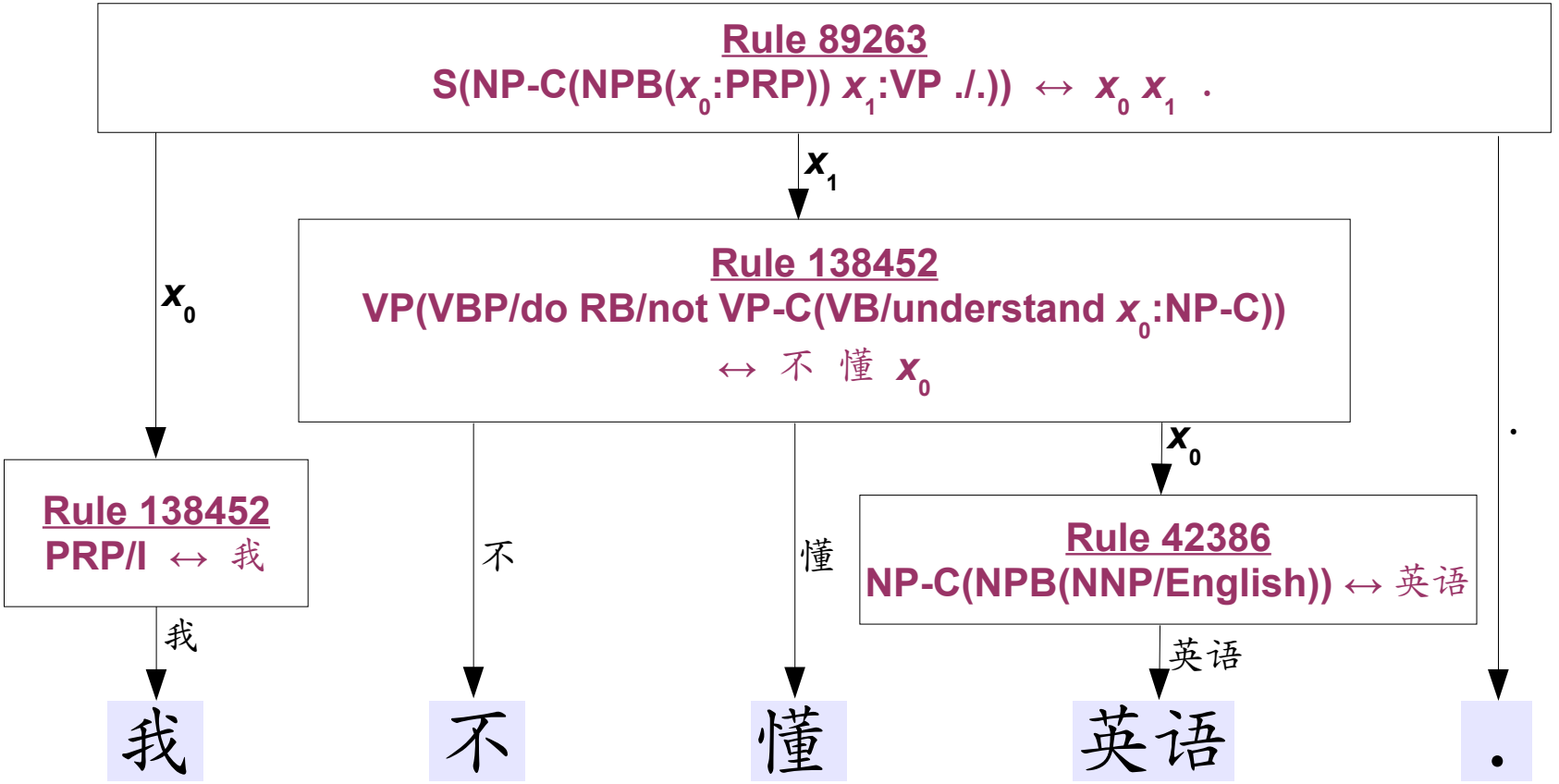
Literally: “I not understand English .”

Decoding Example



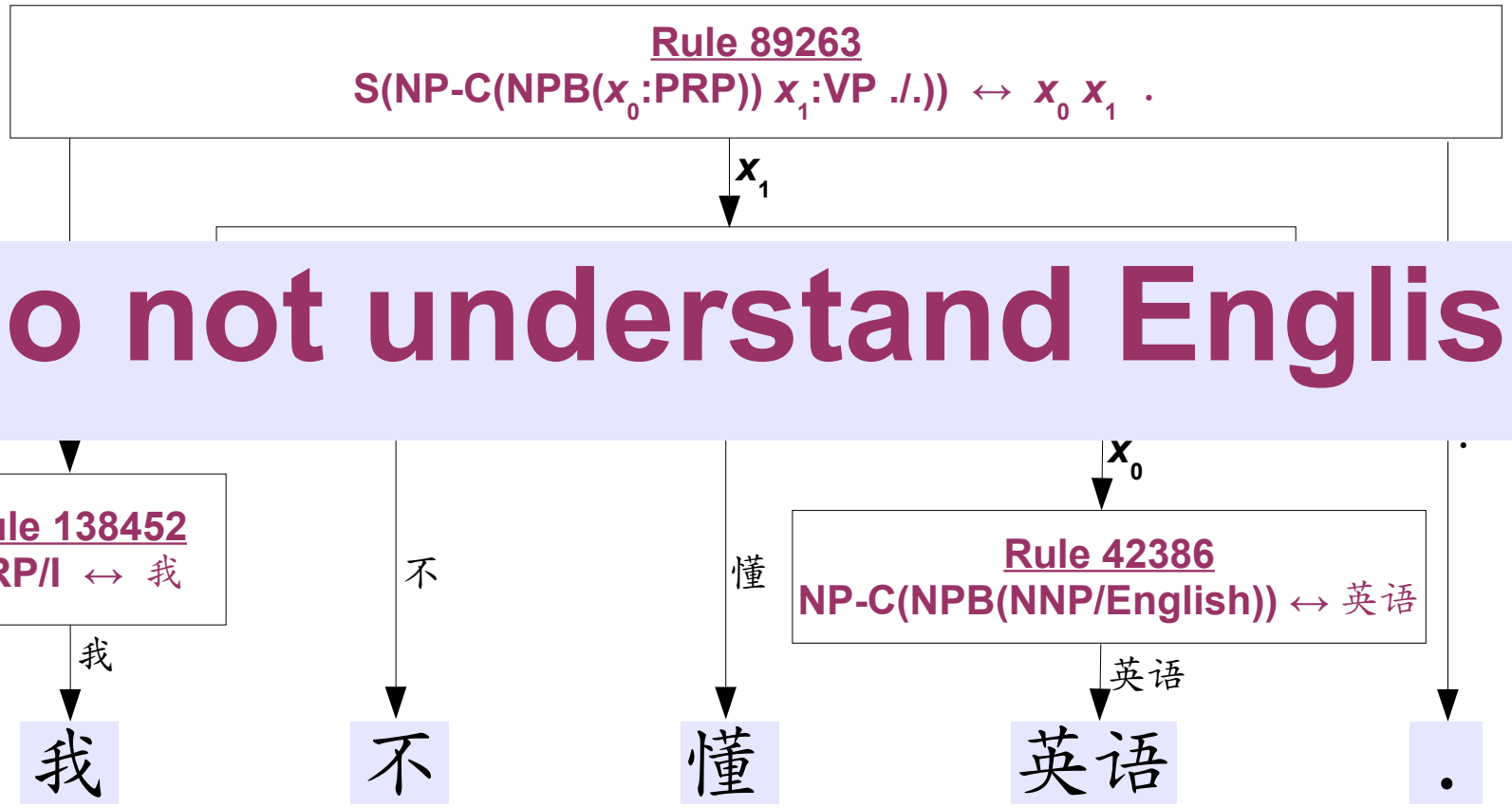
Literally: “I not understand English .”

Decoding Example



Literally: “I not understand English .”

Decoding Example



“I do not understand English .”

Literally: “I not understand English .”

Results: Phrase-based MT

Language	Pre-eval blind test	Evaluation	Post-eval
Arabic	53.79	37.39	**50.16**
Chinese	32.1	33.23	**41.16**
Japanese	44.07	28.31	**33.82**
Korean	35.48	23.74	**30.02**

- **OOPS! Eval scores are very low!**
- **After correcting the phrase tables, scores are more competitive.**

(note: reported numbers are BLEU scores)

Results: Syntax-based MT

Language	Pre-eval blind test	Evaluation	Post-eval
Arabic	43.84	39.62	**44.47**
Chinese	25.73	37.64	**40.08**
Japanese	36.66	27.41	**29.98**
Korean	26.2	25.22	**27.65**

- **Evaluation scores are as expected.**
- **After evaluation, we were able to improve the scores using a trigram LM in search.**

(note: reported numbers are BLEU scores)

Discussion

- Pleasant surprise for Chinese
  - Chinese post-eval syntax-based results were very close to phrase-based results
  - main change: integrating trigram language model into the decoder search
  - this is surprising because the syntax system is currently not learning as many phrase pairs as the phrase-based system

Discussion

- Question Sentences
  - Large percentage of data in this evaluation
  - Syntax for questions is different than the typical “expository text” that our system usually translates.
  - Current parser doesn’t handle questions well.
    - If it did, questions could become a strength rather than a weakness.

Thank You!