Statistical Machine Translation with Weighted Grammars

Matthias Büchse

Chair of Foundations of Programming
Institute of Theoretical Computer Science
Technische Universität Dresden

September 12, 2011

Outline

Outline

Ambiguity in Natural Language

Example (Chiang 2007)

- e = "Australia is one of the few countries that have diplomatic relations with North Korea"
- f = "Aozhou shi yu Beihan you bangjiao de shaoshu guojia zhiyi"

Ambiguity in Natural Language

Example (Chiang 2007)

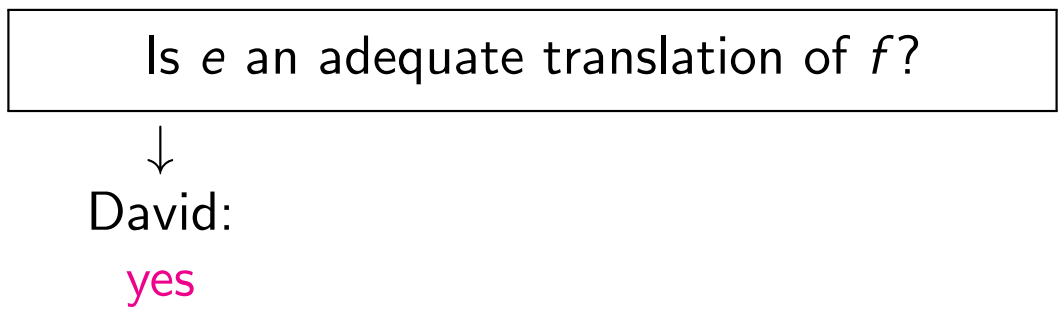
- e = "Australia is one of the few countries that have diplomatic relations with North Korea"
- f = "Aozhou shi yu Beihan you bangjiao de shaoshu guojia zhiyi"

Is *e* an adequate translation of *f*?

Ambiguity in Natural Language

Example (Chiang 2007)

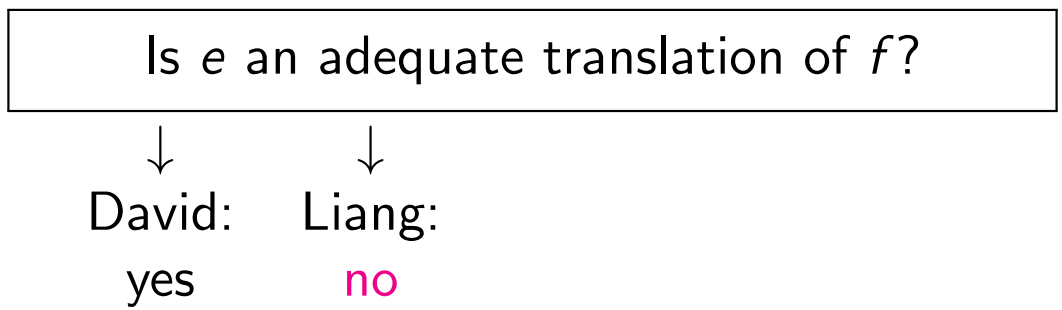
- e = "Australia is one of the few countries that have diplomatic relations with North Korea"
- f = "Aozhou shi yu Beihan you bangjiao de shaoshu guojia zhiyi"



Ambiguity in Natural Language

Example (Chiang 2007)

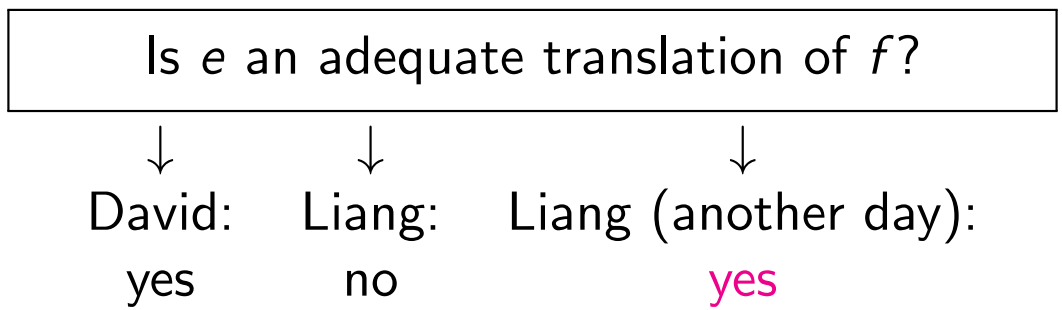
- e = "Australia is one of the few countries that have diplomatic relations with North Korea"
- f = "Aozhou shi yu Beihan you bangjiao de shaoshu guojia zhiyi"



Ambiguity in Natural Language

Example (Chiang 2007)

- e = "Australia is one of the few countries that have diplomatic relations with North Korea"
- f = "Aozhou shi yu Beihan you bangjiao de shaoshu guojia zhiyi"



Ambiguity in Natural Language

Example (Chiang 2007)

- e = "Australia is one of the few countries that have diplomatic relations with North Korea"
- f = "Aozhou shi yu Beihan you bangjiao de shaoshu guojia zhiyi"

What is $P(e \mid f)$?

Ambiguity in Natural Language

Example (Chiang 2007)

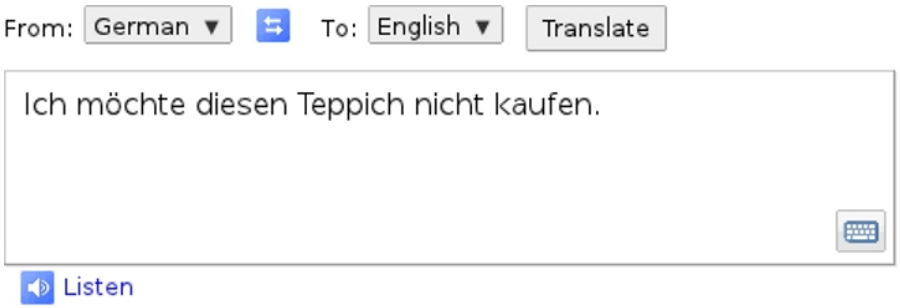
- e = "Australia is one of the few countries that have diplomatic relations with North Korea"
- f = "Aozhou shi yu Beihan you bangjiao de shaoshu guojia zhiyi"

What is $P(e \mid f)$?

$\downarrow$

.42

Search (aka Decoding) in SMT



given $p \in \mathcal{M}(E \mid F)$

Search (aka Decoding) in SMT

given $p \in \mathcal{M}(E \mid F)$ $(\sum_e p(e \mid f) = 1$ for every $f)$

Search (aka Decoding) in SMT

given $p \in \mathcal{M}(E \mid F)$ and $f \in F$,

Search (aka Decoding) in SMT

given $p \in \mathcal{M}(E \mid F)$ and $f \in F$,

determine $\hat{e}$ such that for every $e \in E$:

$$p(\hat{e} \mid f) \geq p(e \mid f) \,. \qquad \text{(not unique)}$$

Search (aka Decoding) in SMT

given $p \in \mathcal{M}(E \mid F)$ and $f \in F$,

determine $\hat{e}$ such that for every $e \in E$:

$$p(\hat{e} \mid f) \geq p(e \mid f) \,. \qquad \text{(not unique)}$$

for short

$$\hat{e} = \operatorname{argmax}_{e \in E} p(e \mid f) \,.$$

Outline

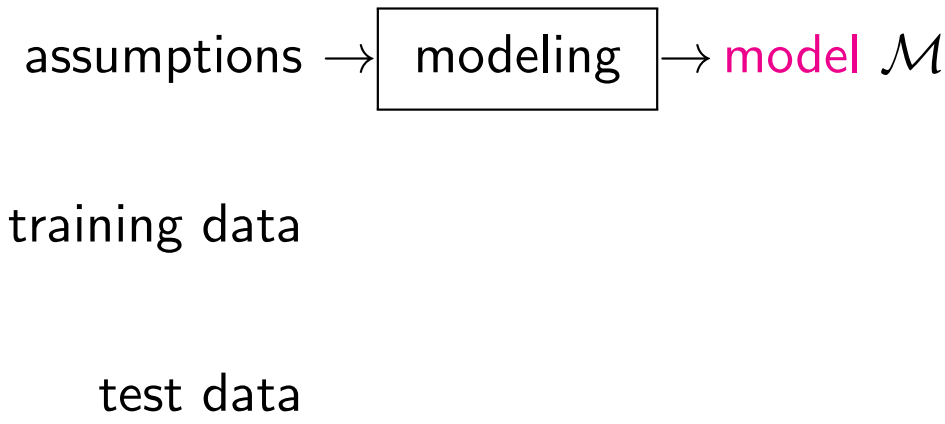
Building an SMT System

assumptions

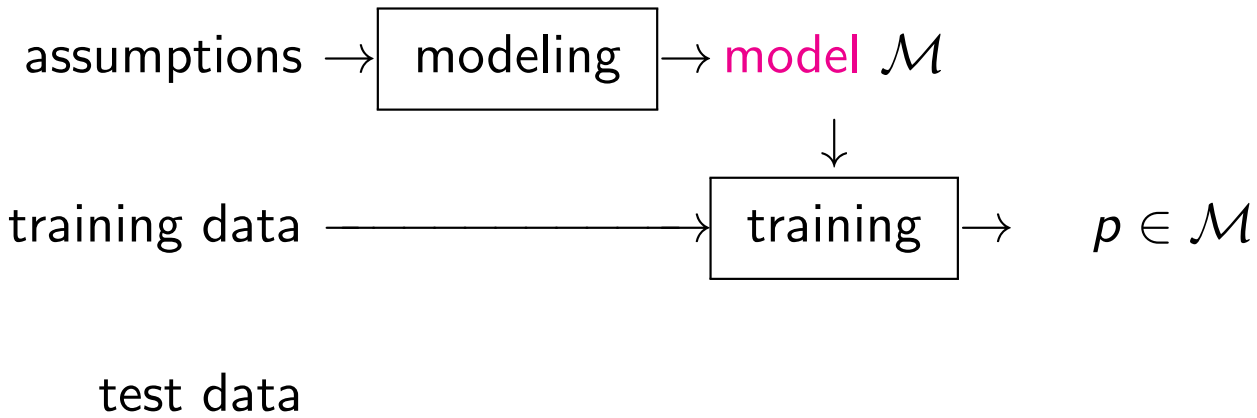
training data

test data

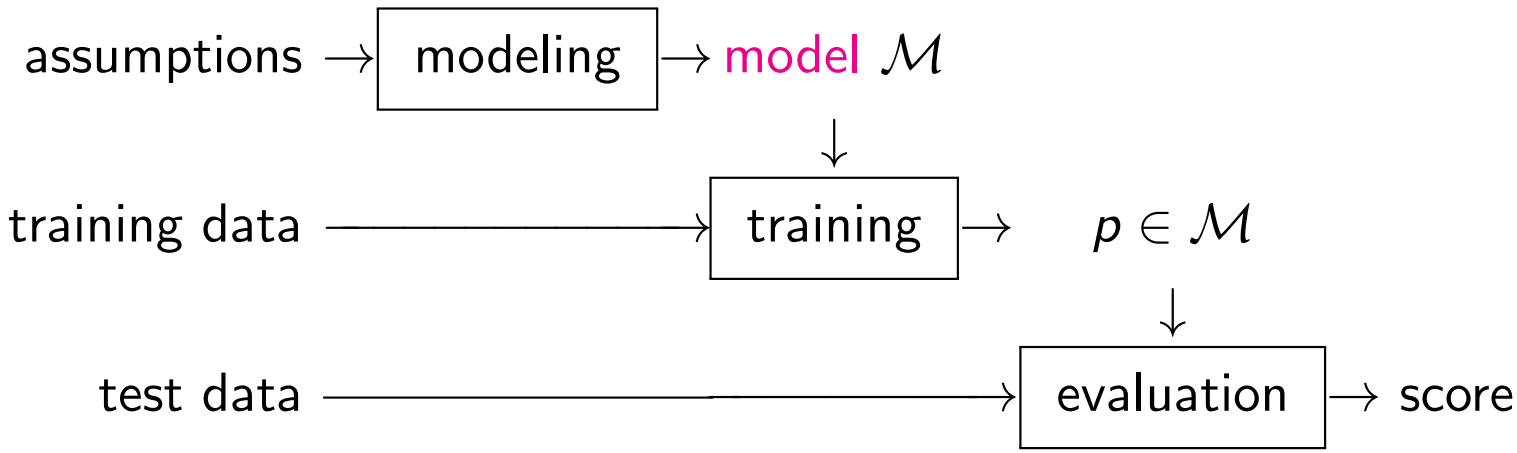
Building an SMT System



Building an SMT System

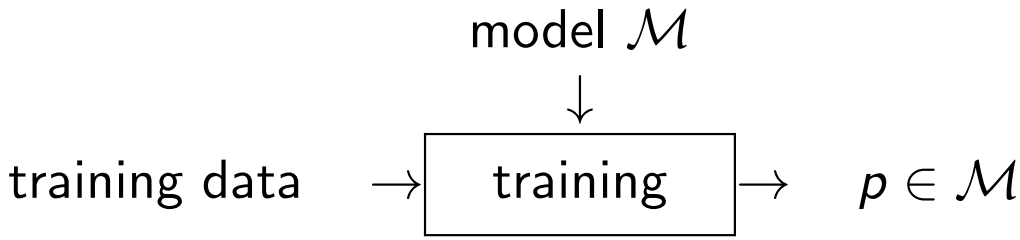


Building an SMT System

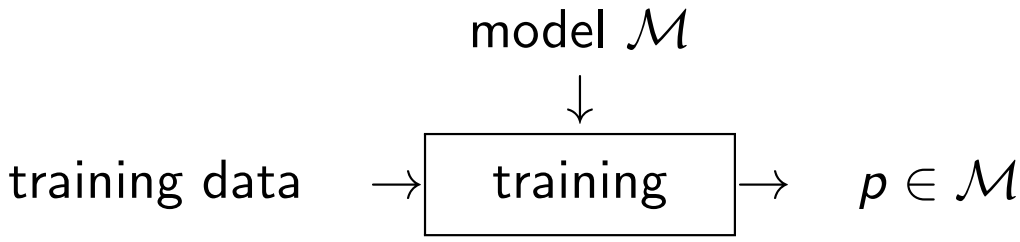


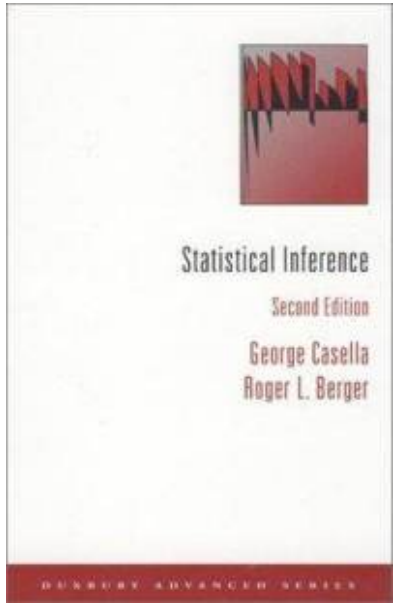
Outline

Training = Point Estimation



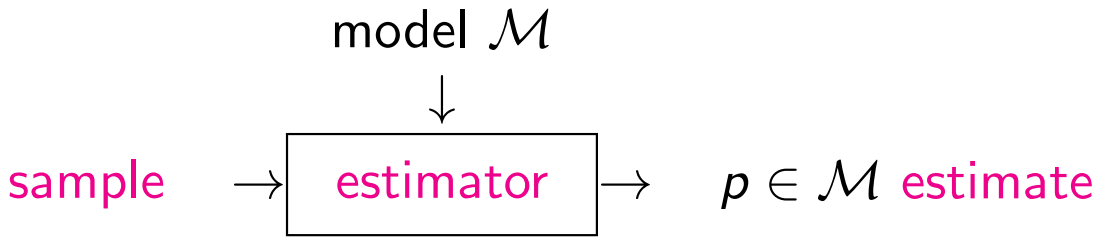
Training = Point Estimation

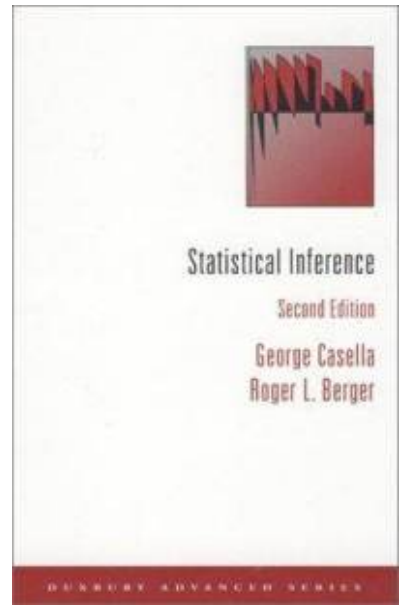




Statistical Inference, Ch. 7: Point Estimation (Casella and Berger 2002)

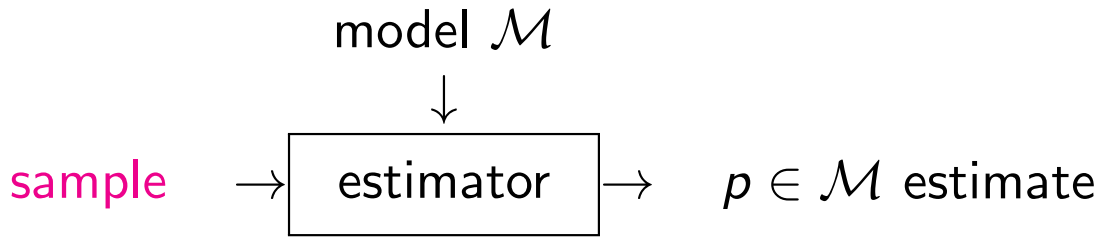
Training = Point Estimation





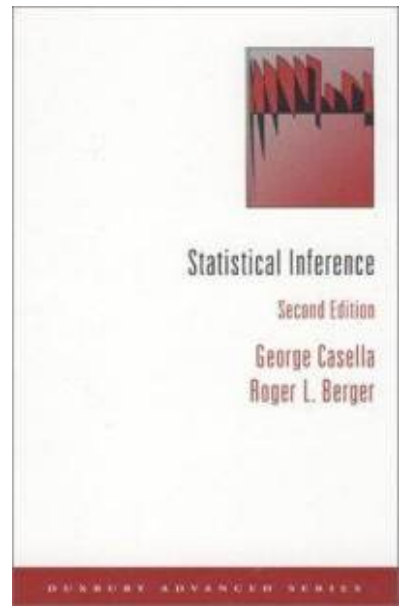
Statistical Inference, Ch. 7: Point Estimation (Casella and Berger 2002)

Training = Point Estimation



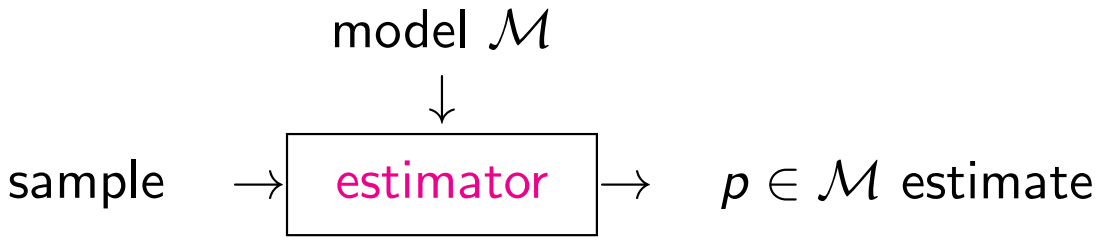
sample

- representative of all translation situations
- e. g., Hong Kong Hansards (parliament proceedings)



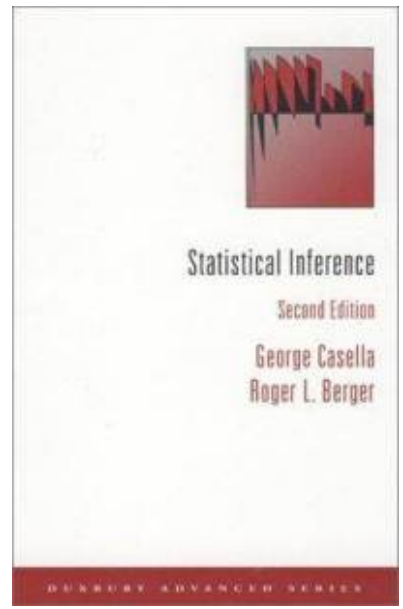
Statistical Inference, Ch. 7: Point Estimation (Casella and Berger 2002)

Training = Point Estimation



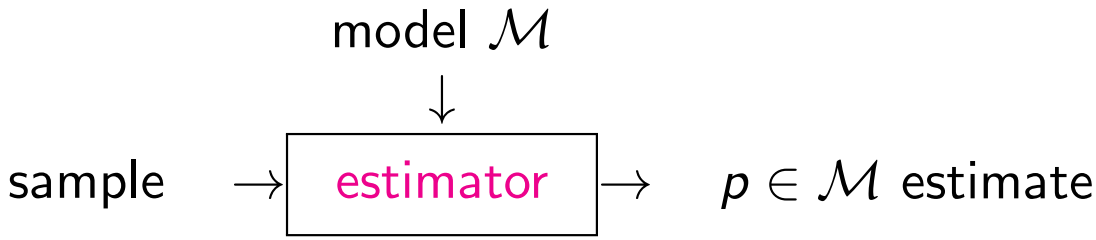
sample
- representative of all translation situations
- e. g., Hong Kong Hansards (parliament proceedings)

estimators
- maximum likelihood
- maximum a-posteriori
- minimum risk
- maximum entropy



Statistical Inference, Ch. 7: Point Estimation (Casella and Berger 2002)

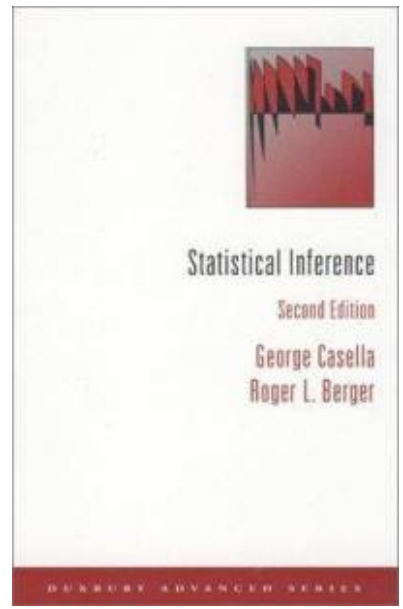
Training = Point Estimation



sample
- representative of all translation situations
- e. g., Hong Kong Hansards (parliament proceedings)

estimators
- maximum likelihood
- maximum a-posteriori
- minimum risk
- maximum entropy

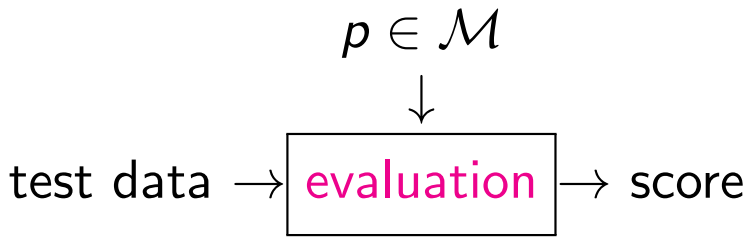
properties
- loss function optimality (Statistical Decision Theory)
- asymptotic optimality (consistency)



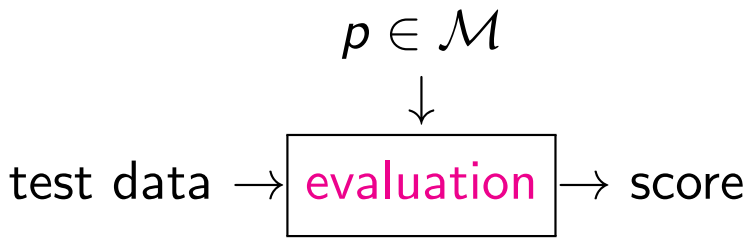
Statistical Inference, Ch. 7: Point Estimation (Casella and Berger 2002)

Outline

Evaluation

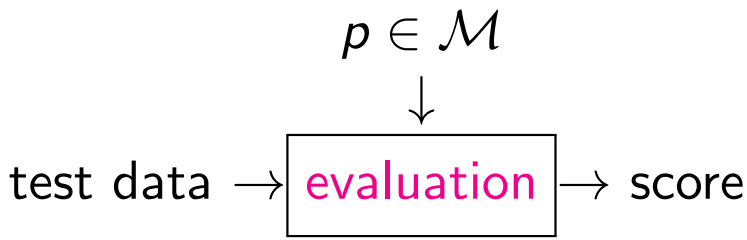


Evaluation



$$\text{training data} \cap \text{test data} = \emptyset$$

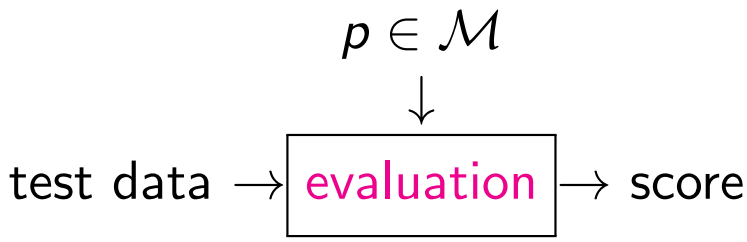
Evaluation



$$\text{training data} \cap \text{test data} = \emptyset$$

method	how it works	goal
BLEU (Papineni et al. 2002)	n-gram precision, brevity penalty	higher values

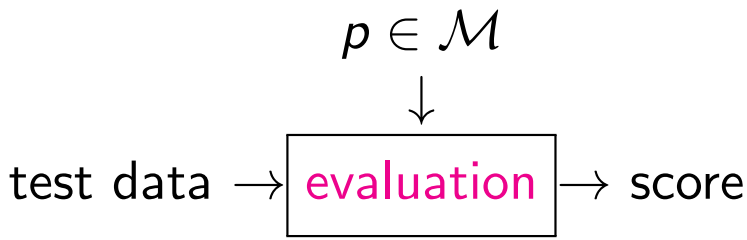
Evaluation



$$\text{training data} \cap \text{test data} = \emptyset$$

method	how it works	goal
BLEU (Papineni et al. 2002)	n-gram precision, brevity penalty	higher values
TER (Snover et al. 2006)	edit distance	lower values

Evaluation

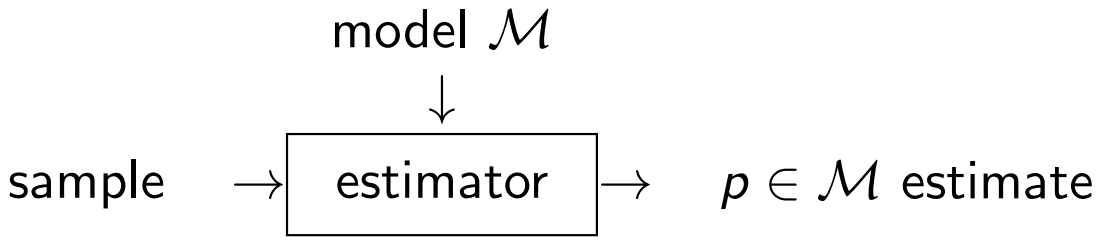


$$\text{training data} \cap \text{test data} = \emptyset$$

method	how it works	goal
BLEU (Papineni et al. 2002)	n-gram precision, brevity penalty	higher values
TER (Snover et al. 2006)	edit distance	lower values
Meteor (Banerjee and Lavie 2005)	alignments, unigram precision, unigram recall, penalty	higher values

Outline

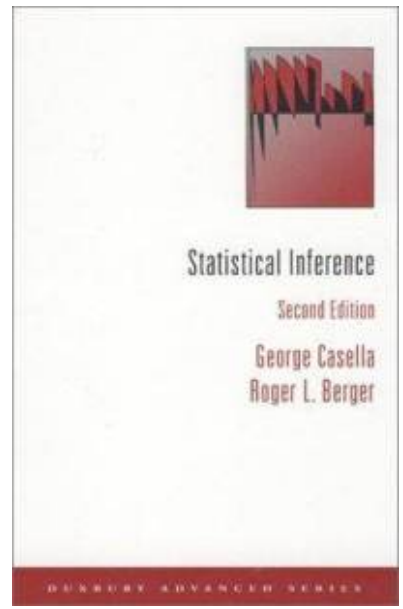
Training = Point Estimation



sample
- representative of all translation situations
- e. g., Hong Kong Hansards (parliament proceedings)

estimators
- maximum likelihood
- maximum a-posteriori
- minimum risk
- maximum entropy

properties
- loss function optimality (Statistical Decision Theory)
- asymptotic optimality (consistency)



Statistical Inference, Ch. 7: Point Estimation (Casella and Berger 2002)

Samples Are Flawed

- data sparsity

Samples Are Flawed

- data sparsity
- noise

"There is no right answer in translation, but there are wrong answers. Translators make a lot of mistakes." Kevin Knight

Samples Are Flawed

- data sparsity
- noise

"There is no right answer in translation, but there are wrong answers. Translators make a lot of mistakes." Kevin Knight

Samples Are Flawed

- data sparsity
- noise

$\Longrightarrow$ overfitting

"There is no right answer in translation, but there are wrong answers. Translators make a lot of mistakes." Kevin Knight

Samples Are Flawed

- data sparsity
- noise

$\Longrightarrow$ overfitting

"There is no right answer in translation, but there are wrong answers. Translators make a lot of mistakes." Kevin Knight

Approaches: {

Samples Are Flawed

- data sparsity
- noise

$\Longrightarrow$ overfitting

"There is no right answer in translation, but there are wrong answers. Translators make a lot of mistakes." Kevin Knight

Approaches:
- use prior (smoothing)

Samples Are Flawed

- data sparsity
- noise

$\Longrightarrow$ overfitting

"There is no right answer in translation, but there are wrong answers. Translators make a lot of mistakes." Kevin Knight

Approaches:
- use prior (smoothing)
- use ad-hoc estimator (heuristics)

Samples Are Flawed

- data sparsity
- noise

$\Longrightarrow$ overfitting

"There is no right answer in translation, but there are wrong answers. Translators make a lot of mistakes." Kevin Knight

Approaches:
- use prior (smoothing)
- use ad-hoc estimator (heuristics)
- model appropriately

Log-linear Model

(Berger, Della Pietra, and Della Pietra 1996; Och and Ney 2002)

- $h_1, \ldots, h_n \colon E \times F \to \mathbb{R}$ features

Log-linear Model

(Berger, Della Pietra, and Della Pietra 1996; Och and Ney 2002)

- $h_1, \ldots, h_n \colon E \times F \to \mathbb{R}$ features
- maximum-entropy formulation:

$$\mathcal{M} = \{p \in \mathcal{M}(E \mid F) \;\; \Big| \;\; \forall i \colon \mathbb{E}_p[h_i] = \mathbb{E}_{\tilde{p}}[h_i]\}$$

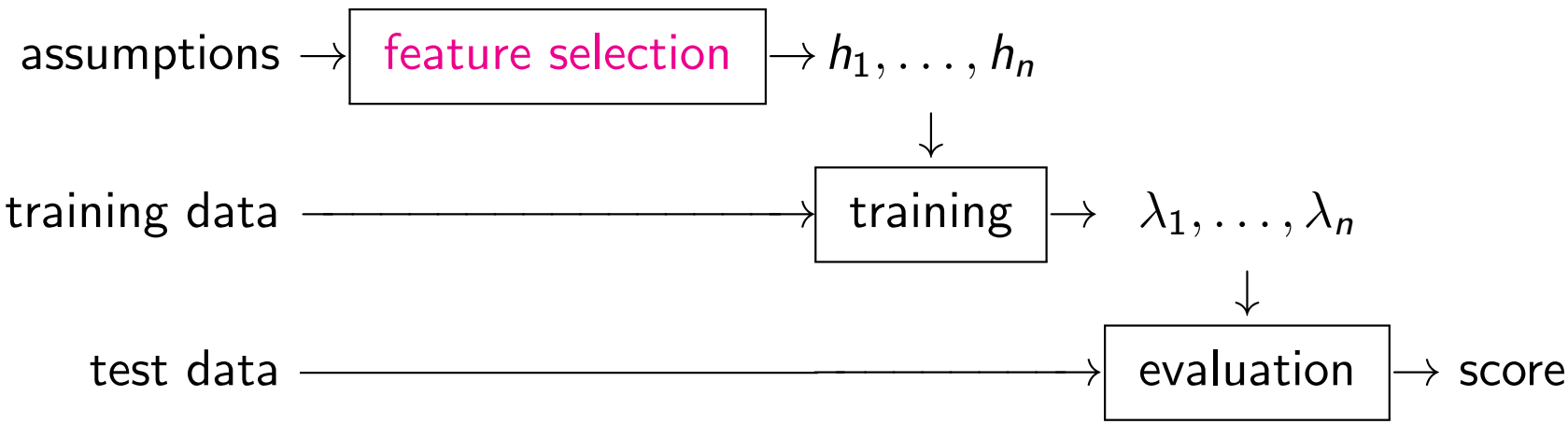
Log-linear Model

(Berger, Della Pietra, and Della Pietra 1996; Och and Ney 2002)

- $h_1, \ldots, h_n \colon E \times F \to \mathbb{R}$ features
- maximum-likelihood formulation:

$$\mathcal{M} = \{p \in \mathcal{M}(E \mid F) \;\Big|\; \exists \lambda_1, \ldots, \lambda_n \in \mathbb{R} \colon p(e \mid f) = \frac{\exp \sum_i \lambda_i \cdot h_i(e, f)}{\sum_{e' \in E} \exp \sum_i \lambda_i \cdot h_i(e', f)}\}$$

Building an SMT System with Log-linear Models



Feature Selection (Examples)

- translation model feature

$$h_1(e, f) = \log p_{\mathrm{TM}}(e \mid f)$$

estimated from a model $\mathcal{M}_{\mathrm{TM}} \subseteq \mathcal{M}(E \mid F)$ via a sample of translations

Feature Selection (Examples)

- translation model feature

$$h_1(e, f) = \log p_{\mathrm{TM}}(e \mid f)$$

estimated from a model $\mathcal{M}_{\mathrm{TM}} \subseteq \mathcal{M}(E \mid F)$ via a sample of translations

- language model feature

$$h_2(e, f) = \log p_{\mathrm{LM}}(e)$$

estimated from a model $\mathcal{M}_{\mathrm{LM}} \subseteq \mathcal{M}(E)$ via a sample of sentences

Feature Selection (Examples)

- translation model feature

$$h_1(e, f) = \log p_{\mathrm{TM}}(e \mid f)$$

  estimated from a model $\mathcal{M}_{\mathrm{TM}} \subseteq \mathcal{M}(E \mid F)$ via a sample of translations

- language model feature

$$h_2(e, f) = \log p_{\mathrm{LM}}(e)$$

  estimated from a model $\mathcal{M}_{\mathrm{LM}} \subseteq \mathcal{M}(E)$ via a sample of sentences

- length feature

$$h_3(e, f) = |e|$$

More Recent Estimators

- minimum-error-rate training (Och 2003)
- minimum-risk annealing (Smith and Eisner 2006)
- large-margin methods (Chiang)

Outline

Example (Chiang 2007)

Recall

- e = "Australia is one of the few countries that have diplomatic relations with North Korea"
- f = "Aozhou shi yu Beihan you bangjiao de shaoshu guojia zhiyi"

Example (Chiang 2007)

Recall

- e = "Australia is one of the few countries that have diplomatic relations with North Korea"
- f = "Aozhou shi yu Beihan you bangjiao de shaoshu guojia zhiyi"

Synchronous Context-Free Grammar G:

π_1: $\mathrm{S} \rightarrow \langle \mathrm{S}\ \mathrm{X},\ \mathrm{S}\ \mathrm{X} \rangle$

π_2: $\mathrm{S} \rightarrow \langle \mathrm{X},\ \mathrm{X} \rangle$

π_3: $\mathrm{X} \rightarrow \langle \text{yu}\ \mathrm{X}_{\boxed{1}}\ \text{you}\ \mathrm{X}_{\boxed{2}},\ \text{have}\ \mathrm{X}_{\boxed{2}}\ \text{with}\ \mathrm{X}_{\boxed{1}} \rangle$

π_4: $\mathrm{X} \rightarrow \langle \mathrm{X}_{\boxed{1}}\ \text{de}\ \mathrm{X}_{\boxed{2}},\ \text{the}\ \mathrm{X}_{\boxed{2}}\ \text{that}\ \mathrm{X}_{\boxed{1}} \rangle$

π_5: $\mathrm{X} \rightarrow \langle \mathrm{X}\ \text{zhiyi},\ \text{one of}\ \mathrm{X} \rangle$

π_6: $\mathrm{X} \rightarrow \langle \text{Aozhou},\ \text{Australia} \rangle$

π_7: $\mathrm{X} \rightarrow \langle \text{Beihan},\ \text{North Korea} \rangle$

π_8: $\mathrm{X} \rightarrow \langle \text{shi},\ \text{is} \rangle$

π_9: $\mathrm{X} \rightarrow \langle \text{bangjiao},\ \text{diplomatic relations} \rangle$

π_{10}: $\mathrm{X} \rightarrow \langle \text{shaoshu guojia},\ \text{few countries} \rangle$

Derivation

$\langle S, S \rangle$

Derivation

$$\langle S,\ S\rangle$$
$$\overset{\pi_1}{\Rightarrow}\ \langle S\ X,\ S\ X\rangle$$

Derivation

$$\langle S,\ S\rangle$$
$$\overset{\pi_1}{\Rightarrow}\ \langle S\ X,\ S\ X\rangle$$

Derivation

$$
\begin{aligned}
& \langle \mathsf{S},\ \mathsf{S} \rangle \\
\overset{\pi_1}{\Rightarrow}\ & \langle \mathsf{S}\ \mathsf{X},\ \mathsf{S}\ \mathsf{X} \rangle \\
\overset{\pi_1}{\Rightarrow}\ & \langle \mathsf{S}\ \mathsf{X}_{\boxed{1}}\ \mathsf{X}_{\boxed{2}},\ \mathsf{S}\ \mathsf{X}_{\boxed{1}}\ \mathsf{X}_{\boxed{2}} \rangle
\end{aligned}
$$

Derivation

$$
\begin{aligned}
& \langle \mathsf{S},\ \mathsf{S} \rangle \\
\overset{\pi_1}{\Rightarrow}\ & \langle \mathsf{S}\ \mathsf{X},\ \mathsf{S}\ \mathsf{X} \rangle \\
\overset{\pi_1}{\Rightarrow}\ & \langle \mathsf{S}\ \mathsf{X}_{\boxed{1}}\ \mathsf{X}_{\boxed{2}},\ \mathsf{S}\ \mathsf{X}_{\boxed{1}}\ \mathsf{X}_{\boxed{2}} \rangle
\end{aligned}
$$

Derivation

$$
\begin{aligned}
& \langle \mathsf{S},\ \mathsf{S} \rangle \\
\overset{\pi_1}{\Rightarrow}\ & \langle \mathsf{S}\ \mathsf{X},\ \mathsf{S}\ \mathsf{X} \rangle \\
\overset{\pi_1}{\Rightarrow}\ & \langle \mathsf{S}\ \mathsf{X}_{\boxed{1}}\ \mathsf{X}_{\boxed{2}},\ \mathsf{S}\ \mathsf{X}_{\boxed{1}}\ \mathsf{X}_{\boxed{2}} \rangle \\
\overset{\pi_2}{\Rightarrow}\ & \langle \mathsf{X}_{\boxed{0}}\ \mathsf{X}_{\boxed{1}}\ \mathsf{X}_{\boxed{2}},\ \mathsf{X}_{\boxed{0}}\ \mathsf{X}_{\boxed{1}}\ \mathsf{X}_{\boxed{2}} \rangle
\end{aligned}
$$

Derivation

$$
\begin{aligned}
& \langle \mathsf{S},\ \mathsf{S} \rangle \\
\overset{\pi_1}{\Rightarrow}\ & \langle \mathsf{S}\ \mathsf{X},\ \mathsf{S}\ \mathsf{X} \rangle \\
\overset{\pi_1}{\Rightarrow}\ & \langle \mathsf{S}\ \mathsf{X}_{\boxed{1}}\ \mathsf{X}_{\boxed{2}},\ \mathsf{S}\ \mathsf{X}_{\boxed{1}}\ \mathsf{X}_{\boxed{2}} \rangle \\
\overset{\pi_2}{\Rightarrow}\ & \langle \mathsf{X}_{\boxed{0}}\ \mathsf{X}_{\boxed{1}}\ \mathsf{X}_{\boxed{2}},\ \mathsf{X}_{\boxed{0}}\ \mathsf{X}_{\boxed{1}}\ \mathsf{X}_{\boxed{2}} \rangle
\end{aligned}
$$

Derivation

$$
\begin{aligned}
& \langle \mathsf{S},\ \mathsf{S} \rangle \\
\overset{\pi_1}{\Rightarrow}\ & \langle \mathsf{S}\ \mathsf{X},\ \mathsf{S}\ \mathsf{X} \rangle \\
\overset{\pi_1}{\Rightarrow}\ & \langle \mathsf{S}\ \mathsf{X}_{\boxed{1}}\ \mathsf{X}_{\boxed{2}},\ \mathsf{S}\ \mathsf{X}_{\boxed{1}}\ \mathsf{X}_{\boxed{2}} \rangle \\
\overset{\pi_2}{\Rightarrow}\ & \langle \mathsf{X}_{\boxed{0}}\ \mathsf{X}_{\boxed{1}}\ \mathsf{X}_{\boxed{2}},\ \mathsf{X}_{\boxed{0}}\ \mathsf{X}_{\boxed{1}}\ \mathsf{X}_{\boxed{2}} \rangle \\
\overset{\pi_6}{\Rightarrow}\ & \langle \text{Aozhou}\ \mathsf{X}_{\boxed{1}}\ \mathsf{X}_{\boxed{2}},\ \text{Australia}\ \mathsf{X}_{\boxed{1}}\ \mathsf{X}_{\boxed{2}} \rangle
\end{aligned}
$$

Derivation

$$
\begin{aligned}
& \langle \mathrm{S},\ \mathrm{S} \rangle \\
\overset{\pi_1}{\Rightarrow}\ & \langle \mathrm{S}\ \mathrm{X},\ \mathrm{S}\ \mathrm{X} \rangle \\
\overset{\pi_1}{\Rightarrow}\ & \langle \mathrm{S}\ \mathrm{X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}},\ \mathrm{S}\ \mathrm{X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}} \rangle \\
\overset{\pi_2}{\Rightarrow}\ & \langle \mathrm{X}_{\boxed{0}}\ \mathrm{X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}},\ \mathrm{X}_{\boxed{0}}\ \mathrm{X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}} \rangle \\
\overset{\pi_6}{\Rightarrow}\ & \langle \mathrm{Aozhou}\ \mathrm{X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}},\ \mathrm{Australia}\ \mathrm{X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}} \rangle
\end{aligned}
$$

Derivation

$$
\begin{aligned}
& \langle \mathrm{S}, \mathrm{S} \rangle \\
\overset{\pi_1}{\Rightarrow} \; & \langle \mathrm{S}\ \mathrm{X}, \mathrm{S}\ \mathrm{X} \rangle \\
\overset{\pi_1}{\Rightarrow} \; & \langle \mathrm{S}\ \mathrm{X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}}, \mathrm{S}\ \mathrm{X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}} \rangle \\
\overset{\pi_2}{\Rightarrow} \; & \langle \mathrm{X}_{\boxed{0}}\ \mathrm{X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}}, \mathrm{X}_{\boxed{0}}\ \mathrm{X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}} \rangle \\
\overset{\pi_6}{\Rightarrow} \; & \langle \text{Aozhou}\ \mathrm{X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}}, \text{Australia}\ \mathrm{X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}} \rangle \\
\overset{\pi_8}{\Rightarrow} \; & \langle \text{Aozhou shi X}, \text{Australia is X} \rangle
\end{aligned}
$$

Derivation

$$
\begin{aligned}
& \langle \mathrm{S},\ \mathrm{S} \rangle \\
\overset{\pi_1}{\Rightarrow}\ & \langle \mathrm{S}\ \mathrm{X},\ \mathrm{S}\ \mathrm{X} \rangle \\
\overset{\pi_1}{\Rightarrow}\ & \langle \mathrm{S}\ \mathrm{X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}},\ \mathrm{S}\ \mathrm{X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}} \rangle \\
\overset{\pi_2}{\Rightarrow}\ & \langle \mathrm{X}_{\boxed{0}}\ \mathrm{X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}},\ \mathrm{X}_{\boxed{0}}\ \mathrm{X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}} \rangle \\
\overset{\pi_6}{\Rightarrow}\ & \langle \text{Aozhou}\ \mathrm{X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}},\ \text{Australia}\ \mathrm{X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}} \rangle \\
\overset{\pi_8}{\Rightarrow}\ & \langle \text{Aozhou shi X},\ \text{Australia is X} \rangle
\end{aligned}
$$

Derivation

$$
\begin{aligned}
& \langle \mathrm{S},\ \mathrm{S} \rangle \\
\overset{\pi_1}{\Rightarrow}\ & \langle \mathrm{S\ X},\ \mathrm{S\ X} \rangle \\
\overset{\pi_1}{\Rightarrow}\ & \langle \mathrm{S\ X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}},\ \mathrm{S\ X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}} \rangle \\
\overset{\pi_2}{\Rightarrow}\ & \langle \mathrm{X}_{\boxed{0}}\ \mathrm{X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}},\ \mathrm{X}_{\boxed{0}}\ \mathrm{X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}} \rangle \\
\overset{\pi_6}{\Rightarrow}\ & \langle \mathrm{Aozhou\ X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}},\ \mathrm{Australia\ X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}} \rangle \\
\overset{\pi_8}{\Rightarrow}\ & \langle \mathrm{Aozhou\ shi\ X},\ \mathrm{Australia\ is\ X} \rangle \\
\overset{\pi_5}{\Rightarrow}\ & \langle \mathrm{Aozhou\ shi\ X\ zhiyi},\ \mathrm{Australia\ is\ one\ of\ X} \rangle
\end{aligned}
$$

Derivation

$$
\begin{aligned}
& \langle \mathrm{S},\ \mathrm{S} \rangle \\
\overset{\pi_1}{\Rightarrow}\ & \langle \mathrm{S}\ \mathrm{X},\ \mathrm{S}\ \mathrm{X} \rangle \\
\overset{\pi_1}{\Rightarrow}\ & \langle \mathrm{S}\ \mathrm{X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}},\ \mathrm{S}\ \mathrm{X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}} \rangle \\
\overset{\pi_2}{\Rightarrow}\ & \langle \mathrm{X}_{\boxed{0}}\ \mathrm{X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}},\ \mathrm{X}_{\boxed{0}}\ \mathrm{X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}} \rangle \\
\overset{\pi_6}{\Rightarrow}\ & \langle \text{Aozhou}\ \mathrm{X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}},\ \text{Australia}\ \mathrm{X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}} \rangle \\
\overset{\pi_8}{\Rightarrow}\ & \langle \text{Aozhou shi X},\ \text{Australia is X} \rangle \\
\overset{\pi_5}{\Rightarrow}\ & \langle \text{Aozhou shi X zhiyi},\ \text{Australia is one of X} \rangle
\end{aligned}
$$

Derivation

$$
\begin{aligned}
& \langle \mathrm{S},\ \mathrm{S} \rangle \\
\overset{\pi_1}{\Rightarrow}\ & \langle \mathrm{S\ X},\ \mathrm{S\ X} \rangle \\
\overset{\pi_1}{\Rightarrow}\ & \langle \mathrm{S\ X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}},\ \mathrm{S\ X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}} \rangle \\
\overset{\pi_2}{\Rightarrow}\ & \langle \mathrm{X}_{\boxed{0}}\ \mathrm{X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}},\ \mathrm{X}_{\boxed{0}}\ \mathrm{X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}} \rangle \\
\overset{\pi_6}{\Rightarrow}\ & \langle \mathrm{Aozhou\ X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}},\ \mathrm{Australia\ X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}} \rangle \\
\overset{\pi_8}{\Rightarrow}\ & \langle \mathrm{Aozhou\ shi\ X},\ \mathrm{Australia\ is\ X} \rangle \\
\overset{\pi_5}{\Rightarrow}\ & \langle \mathrm{Aozhou\ shi\ X\ zhiyi},\ \mathrm{Australia\ is\ one\ of\ X} \rangle \\
\overset{\pi_4}{\Rightarrow}\ & \langle \mathrm{Aozhou\ shi\ X}_{\boxed{1}}\ \mathrm{de\ X}_{\boxed{2}}\ \mathrm{zhiyi}, \\
& \ \mathrm{Australia\ is\ one\ of\ the\ X}_{\boxed{2}}\ \mathrm{that\ X}_{\boxed{1}} \rangle
\end{aligned}
$$

Derivation

$$
\begin{aligned}
& \langle \mathrm{S},\ \mathrm{S} \rangle \\
\overset{\pi_1}{\Rightarrow}\ & \langle \mathrm{S\ X},\ \mathrm{S\ X} \rangle \\
\overset{\pi_1}{\Rightarrow}\ & \langle \mathrm{S\ X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}},\ \mathrm{S\ X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}} \rangle \\
\overset{\pi_2}{\Rightarrow}\ & \langle \mathrm{X}_{\boxed{0}}\ \mathrm{X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}},\ \mathrm{X}_{\boxed{0}}\ \mathrm{X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}} \rangle \\
\overset{\pi_6}{\Rightarrow}\ & \langle \mathrm{Aozhou\ X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}},\ \mathrm{Australia\ X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}} \rangle \\
\overset{\pi_8}{\Rightarrow}\ & \langle \mathrm{Aozhou\ shi\ X},\ \mathrm{Australia\ is\ X} \rangle \\
\overset{\pi_5}{\Rightarrow}\ & \langle \mathrm{Aozhou\ shi\ X\ zhiyi},\ \mathrm{Australia\ is\ one\ of\ X} \rangle \\
\overset{\pi_4}{\Rightarrow}\ & \langle \mathrm{Aozhou\ shi\ X}_{\boxed{1}}\ \mathrm{de\ X}_{\boxed{2}}\ \mathrm{zhiyi}, \\
& \quad \mathrm{Australia\ is\ one\ of\ the\ X}_{\boxed{2}}\ \mathrm{that\ X}_{\boxed{1}} \rangle
\end{aligned}
$$

Derivation

$$
\begin{aligned}
& \langle \mathrm{S},\ \mathrm{S} \rangle \\
\overset{\pi_1}{\Rightarrow}\ & \langle \mathrm{S\ X},\ \mathrm{S\ X} \rangle \\
\overset{\pi_1}{\Rightarrow}\ & \langle \mathrm{S\ X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}},\ \mathrm{S\ X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}} \rangle \\
\overset{\pi_2}{\Rightarrow}\ & \langle \mathrm{X}_{\boxed{0}}\ \mathrm{X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}},\ \mathrm{X}_{\boxed{0}}\ \mathrm{X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}} \rangle \\
\overset{\pi_6}{\Rightarrow}\ & \langle \mathrm{Aozhou\ X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}},\ \mathrm{Australia\ X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}} \rangle \\
\overset{\pi_8}{\Rightarrow}\ & \langle \mathrm{Aozhou\ shi\ X},\ \mathrm{Australia\ is\ X} \rangle \\
\overset{\pi_5}{\Rightarrow}\ & \langle \mathrm{Aozhou\ shi\ X\ zhiyi},\ \mathrm{Australia\ is\ one\ of\ X} \rangle \\
\overset{\pi_4}{\Rightarrow}\ & \langle \mathrm{Aozhou\ shi\ X}_{\boxed{1}}\ \mathrm{de\ X}_{\boxed{2}}\ \mathrm{zhiyi}, \\
& \quad \mathrm{Australia\ is\ one\ of\ the\ X}_{\boxed{2}}\ \mathrm{that\ X}_{\boxed{1}} \rangle \\
\overset{\pi_3}{\Rightarrow}\ & \langle \mathrm{Aozhou\ shi\ yu\ X}_{\boxed{1}}\ \mathrm{you\ X}_{\boxed{0}}\ \mathrm{de\ X}_{\boxed{2}}\ \mathrm{zhiyi}, \\
& \quad \mathrm{Australia\ is\ one\ of\ the\ X}_{\boxed{2}}\ \mathrm{that\ have\ X}_{\boxed{0}}\ \mathrm{with\ X}_{\boxed{1}} \rangle
\end{aligned}
$$

Derivation

$$
\begin{aligned}
& \langle \mathrm{S},\ \mathrm{S} \rangle \\
\overset{\pi_1}{\Rightarrow}\ & \langle \mathrm{S\ X},\ \mathrm{S\ X} \rangle \\
\overset{\pi_1}{\Rightarrow}\ & \langle \mathrm{S\ X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}},\ \mathrm{S\ X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}} \rangle \\
\overset{\pi_2}{\Rightarrow}\ & \langle \mathrm{X}_{\boxed{0}}\ \mathrm{X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}},\ \mathrm{X}_{\boxed{0}}\ \mathrm{X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}} \rangle \\
\overset{\pi_6}{\Rightarrow}\ & \langle \mathrm{Aozhou\ X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}},\ \mathrm{Australia\ X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}} \rangle \\
\overset{\pi_8}{\Rightarrow}\ & \langle \mathrm{Aozhou\ shi\ X},\ \mathrm{Australia\ is\ X} \rangle \\
\overset{\pi_5}{\Rightarrow}\ & \langle \mathrm{Aozhou\ shi\ X\ zhiyi},\ \mathrm{Australia\ is\ one\ of\ X} \rangle \\
\overset{\pi_4}{\Rightarrow}\ & \langle \mathrm{Aozhou\ shi\ X}_{\boxed{1}}\ \mathrm{de\ X}_{\boxed{2}}\ \mathrm{zhiyi}, \\
& \quad \mathrm{Australia\ is\ one\ of\ the\ X}_{\boxed{2}}\ \mathrm{that\ X}_{\boxed{1}} \rangle \\
\overset{\pi_3}{\Rightarrow}\ & \langle \mathrm{Aozhou\ shi\ yu\ X}_{\boxed{1}}\ \mathrm{you\ X}_{\boxed{0}}\ \mathrm{de\ X}_{\boxed{2}}\ \mathrm{zhiyi}, \\
& \quad \mathrm{Australia\ is\ one\ of\ the\ X}_{\boxed{2}}\ \mathrm{that\ have\ X}_{\boxed{0}}\ \mathrm{with\ X}_{\boxed{1}} \rangle
\end{aligned}
$$

Derivation

$$
\begin{aligned}
& \langle \mathrm{S},\ \mathrm{S} \rangle \\
\overset{\pi_1}{\Rightarrow}\ & \langle \mathrm{S\ X},\ \mathrm{S\ X} \rangle \\
\overset{\pi_1}{\Rightarrow}\ & \langle \mathrm{S\ X_{\boxed{1}}\ X_{\boxed{2}}},\ \mathrm{S\ X_{\boxed{1}}\ X_{\boxed{2}}} \rangle \\
\overset{\pi_2}{\Rightarrow}\ & \langle \mathrm{X_{\boxed{0}}\ X_{\boxed{1}}\ X_{\boxed{2}}},\ \mathrm{X_{\boxed{0}}\ X_{\boxed{1}}\ X_{\boxed{2}}} \rangle \\
\overset{\pi_6}{\Rightarrow}\ & \langle \mathrm{Aozhou\ X_{\boxed{1}}\ X_{\boxed{2}}},\ \mathrm{Australia\ X_{\boxed{1}}\ X_{\boxed{2}}} \rangle \\
\overset{\pi_8}{\Rightarrow}\ & \langle \mathrm{Aozhou\ shi\ X},\ \mathrm{Australia\ is\ X} \rangle \\
\overset{\pi_5}{\Rightarrow}\ & \langle \mathrm{Aozhou\ shi\ X\ zhiyi},\ \mathrm{Australia\ is\ one\ of\ X} \rangle \\
\overset{\pi_4}{\Rightarrow}\ & \langle \mathrm{Aozhou\ shi\ X_{\boxed{1}}\ de\ X_{\boxed{2}}\ zhiyi}, \\
& \quad \mathrm{Australia\ is\ one\ of\ the\ X_{\boxed{2}}\ that\ X_{\boxed{1}}} \rangle \\
\overset{\pi_3}{\Rightarrow}\ & \langle \mathrm{Aozhou\ shi\ yu\ X_{\boxed{1}}\ you\ X_{\boxed{0}}\ de\ X_{\boxed{2}}\ zhiyi}, \\
& \quad \mathrm{Australia\ is\ one\ of\ the\ X_{\boxed{2}}\ that\ have\ X_{\boxed{0}}\ with\ X_{\boxed{1}}} \rangle \\
\overset{\pi_7}{\Rightarrow}\ & \langle \mathrm{Aozhou\ shi\ yu\ Beihan\ you\ X_{\boxed{0}}\ de\ X_{\boxed{2}}\ zhiyi}, \\
& \quad \mathrm{Australia\ is\ one\ of\ the\ X_{\boxed{2}}\ that\ have\ X_{\boxed{0}}\ with\ North\ Korea} \rangle
\end{aligned}
$$

Derivation

$$
\begin{aligned}
& \langle \mathrm{S},\ \mathrm{S} \rangle \\
\overset{\pi_1}{\Rightarrow}\ & \langle \mathrm{S\ X},\ \mathrm{S\ X} \rangle \\
\overset{\pi_1}{\Rightarrow}\ & \langle \mathrm{S\ X_{\boxed{1}}\ X_{\boxed{2}}},\ \mathrm{S\ X_{\boxed{1}}\ X_{\boxed{2}}} \rangle \\
\overset{\pi_2}{\Rightarrow}\ & \langle \mathrm{X_{\boxed{0}}\ X_{\boxed{1}}\ X_{\boxed{2}}},\ \mathrm{X_{\boxed{0}}\ X_{\boxed{1}}\ X_{\boxed{2}}} \rangle \\
\overset{\pi_6}{\Rightarrow}\ & \langle \mathrm{Aozhou\ X_{\boxed{1}}\ X_{\boxed{2}}},\ \mathrm{Australia\ X_{\boxed{1}}\ X_{\boxed{2}}} \rangle \\
\overset{\pi_8}{\Rightarrow}\ & \langle \mathrm{Aozhou\ shi\ X},\ \mathrm{Australia\ is\ X} \rangle \\
\overset{\pi_5}{\Rightarrow}\ & \langle \mathrm{Aozhou\ shi\ X\ zhiyi},\ \mathrm{Australia\ is\ one\ of\ X} \rangle \\
\overset{\pi_4}{\Rightarrow}\ & \langle \mathrm{Aozhou\ shi\ X_{\boxed{1}}\ de\ X_{\boxed{2}}\ zhiyi}, \\
& \quad \mathrm{Australia\ is\ one\ of\ the\ X_{\boxed{2}}\ that\ X_{\boxed{1}}} \rangle \\
\overset{\pi_3}{\Rightarrow}\ & \langle \mathrm{Aozhou\ shi\ yu\ X_{\boxed{1}}\ you\ X_{\boxed{0}}\ de\ X_{\boxed{2}}\ zhiyi}, \\
& \quad \mathrm{Australia\ is\ one\ of\ the\ X_{\boxed{2}}\ that\ have\ X_{\boxed{0}}\ with\ X_{\boxed{1}}} \rangle \\
\overset{\pi_7}{\Rightarrow}\ & \langle \mathrm{Aozhou\ shi\ yu\ Beihan\ you\ X_{\boxed{0}}\ de\ X_{\boxed{2}}\ zhiyi}, \\
& \quad \mathrm{Australia\ is\ one\ of\ the\ X_{\boxed{2}}\ that\ have\ X_{\boxed{0}}\ with\ North\ Korea} \rangle
\end{aligned}
$$

Derivation

$$
\begin{aligned}
& \langle \mathrm{S},\ \mathrm{S} \rangle \\
\overset{\pi_1}{\Rightarrow}\ & \langle \mathrm{S\ X},\ \mathrm{S\ X} \rangle \\
\overset{\pi_1}{\Rightarrow}\ & \langle \mathrm{S\ X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}},\ \mathrm{S\ X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}} \rangle \\
\overset{\pi_2}{\Rightarrow}\ & \langle \mathrm{X}_{\boxed{0}}\ \mathrm{X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}},\ \mathrm{X}_{\boxed{0}}\ \mathrm{X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}} \rangle \\
\overset{\pi_6}{\Rightarrow}\ & \langle \mathrm{Aozhou\ X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}},\ \mathrm{Australia\ X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}} \rangle \\
\overset{\pi_8}{\Rightarrow}\ & \langle \mathrm{Aozhou\ shi\ X},\ \mathrm{Australia\ is\ X} \rangle \\
\overset{\pi_5}{\Rightarrow}\ & \langle \mathrm{Aozhou\ shi\ X\ zhiyi},\ \mathrm{Australia\ is\ one\ of\ X} \rangle \\
\overset{\pi_4}{\Rightarrow}\ & \langle \mathrm{Aozhou\ shi\ X}_{\boxed{1}}\ \mathrm{de\ X}_{\boxed{2}}\ \mathrm{zhiyi}, \\
& \quad \mathrm{Australia\ is\ one\ of\ the\ X}_{\boxed{2}}\ \mathrm{that\ X}_{\boxed{1}} \rangle \\
\overset{\pi_3}{\Rightarrow}\ & \langle \mathrm{Aozhou\ shi\ yu\ X}_{\boxed{1}}\ \mathrm{you\ X}_{\boxed{0}}\ \mathrm{de\ X}_{\boxed{2}}\ \mathrm{zhiyi}, \\
& \quad \mathrm{Australia\ is\ one\ of\ the\ X}_{\boxed{2}}\ \mathrm{that\ have\ X}_{\boxed{0}}\ \mathrm{with\ X}_{\boxed{1}} \rangle \\
\overset{\pi_7}{\Rightarrow}\ & \langle \mathrm{Aozhou\ shi\ yu\ Beihan\ you\ X}_{\boxed{0}}\ \mathrm{de\ X}_{\boxed{2}}\ \mathrm{zhiyi}, \\
& \quad \mathrm{Australia\ is\ one\ of\ the\ X}_{\boxed{2}}\ \mathrm{that\ have\ X}_{\boxed{0}}\ \mathrm{with\ North\ Korea} \rangle \\
\overset{\pi_9}{\Rightarrow}\ & \langle \mathrm{Aozhou\ shi\ yu\ Beihan\ you\ bangjiao\ de\ X}_{\boxed{2}}\ \mathrm{zhiyi}, \\
& \quad \mathrm{Australia\ is\ one\ of\ the\ X}_{\boxed{2}}\ \mathrm{that\ have\ diplomatic\ relations\ with\ North\ Korea} \rangle
\end{aligned}
$$

Derivation

$\langle S, S \rangle$

$\overset{\pi_1}{\Rightarrow} \langle S\ X, S\ X \rangle$

$\overset{\pi_1}{\Rightarrow} \langle S\ X_{\boxed{1}}\ X_{\boxed{2}}, S\ X_{\boxed{1}}\ X_{\boxed{2}} \rangle$

$\overset{\pi_2}{\Rightarrow} \langle X_{\boxed{0}}\ X_{\boxed{1}}\ X_{\boxed{2}}, X_{\boxed{0}}\ X_{\boxed{1}}\ X_{\boxed{2}} \rangle$

$\overset{\pi_6}{\Rightarrow} \langle \text{Aozhou } X_{\boxed{1}}\ X_{\boxed{2}}, \text{Australia } X_{\boxed{1}}\ X_{\boxed{2}} \rangle$

$\overset{\pi_8}{\Rightarrow} \langle \text{Aozhou shi X, Australia is X} \rangle$

$\overset{\pi_5}{\Rightarrow} \langle \text{Aozhou shi X zhiyi, Australia is one of X} \rangle$

$\overset{\pi_4}{\Rightarrow} \langle \text{Aozhou shi } X_{\boxed{1}} \text{ de } X_{\boxed{2}} \text{ zhiyi,}$
$\text{Australia is one of the } X_{\boxed{2}} \text{ that } X_{\boxed{1}} \rangle$

$\overset{\pi_3}{\Rightarrow} \langle \text{Aozhou shi yu } X_{\boxed{1}} \text{ you } X_{\boxed{0}} \text{ de } X_{\boxed{2}} \text{ zhiyi,}$
$\text{Australia is one of the } X_{\boxed{2}} \text{ that have } X_{\boxed{0}} \text{ with } X_{\boxed{1}} \rangle$

$\overset{\pi_7}{\Rightarrow} \langle \text{Aozhou shi yu Beihan you } X_{\boxed{0}} \text{ de } X_{\boxed{2}} \text{ zhiyi,}$
$\text{Australia is one of the } X_{\boxed{2}} \text{ that have } X_{\boxed{0}} \text{ with North Korea} \rangle$

$\overset{\pi_9}{\Rightarrow} \langle \text{Aozhou shi yu Beihan you bangjiao de } X_{\boxed{2}} \text{ zhiyi,}$
$\text{Australia is one of the } X_{\boxed{2}} \text{ that have diplomatic relations with North Korea} \rangle$

Derivation

$\langle$S, S$\rangle$

$\overset{\pi_1}{\Rightarrow}$ $\langle$S X, S X$\rangle$

$\overset{\pi_1}{\Rightarrow}$ $\langle$S $X_{\boxed{1}}$ $X_{\boxed{2}}$, S $X_{\boxed{1}}$ $X_{\boxed{2}}$$\rangle$

$\overset{\pi_2}{\Rightarrow}$ $\langle$$X_{\boxed{0}}$ $X_{\boxed{1}}$ $X_{\boxed{2}}$, $X_{\boxed{0}}$ $X_{\boxed{1}}$ $X_{\boxed{2}}$$\rangle$

$\overset{\pi_6}{\Rightarrow}$ $\langle$Aozhou $X_{\boxed{1}}$ $X_{\boxed{2}}$, Australia $X_{\boxed{1}}$ $X_{\boxed{2}}$$\rangle$

$\overset{\pi_8}{\Rightarrow}$ $\langle$Aozhou shi X, Australia is X$\rangle$

$\overset{\pi_5}{\Rightarrow}$ $\langle$Aozhou shi X zhiyi, Australia is one of X$\rangle$

$\overset{\pi_4}{\Rightarrow}$ $\langle$Aozhou shi $X_{\boxed{1}}$ de $X_{\boxed{2}}$ zhiyi,
Australia is one of the $X_{\boxed{2}}$ that $X_{\boxed{1}}$$\rangle$

$\overset{\pi_3}{\Rightarrow}$ $\langle$Aozhou shi yu $X_{\boxed{1}}$ you $X_{\boxed{0}}$ de $X_{\boxed{2}}$ zhiyi,
Australia is one of the $X_{\boxed{2}}$ that have $X_{\boxed{0}}$ with $X_{\boxed{1}}$$\rangle$

$\overset{\pi_7}{\Rightarrow}$ $\langle$Aozhou shi yu Beihan you $X_{\boxed{0}}$ de $X_{\boxed{2}}$ zhiyi,
Australia is one of the $X_{\boxed{2}}$ that have $X_{\boxed{0}}$ with North Korea$\rangle$

$\overset{\pi_9}{\Rightarrow}$ $\langle$Aozhou shi yu Beihan you bangjiao de $X_{\boxed{2}}$ zhiyi,
Australia is one of the $X_{\boxed{2}}$ that have diplomatic relations with North Korea$\rangle$

$\overset{\pi_{10}}{\Rightarrow}$ $\langle$Aozhou shi yu Beihan you bangjiao de shaoshu guojia zhiyi,
Australia is one of the few countries that have diplomatic relations with N. K.$\rangle$

Derivation

$\langle$S, S$\rangle$

$\stackrel{\pi_1}{\Rightarrow}$ $\langle$S X, S X$\rangle$

$\stackrel{\pi_1}{\Rightarrow}$ $\langle$S $X_{\boxed{1}}$ $X_{\boxed{2}}$, S $X_{\boxed{1}}$ $X_{\boxed{2}}$$\rangle$

$\stackrel{\pi_2}{\Rightarrow}$ $\langle$$X_{\boxed{0}}$ $X_{\boxed{1}}$ $X_{\boxed{2}}$, $X_{\boxed{0}}$ $X_{\boxed{1}}$ $X_{\boxed{2}}$$\rangle$

$\stackrel{\pi_6}{\Rightarrow}$ $\langle$Aozhou $X_{\boxed{1}}$ $X_{\boxed{2}}$, Australia $X_{\boxed{1}}$ $X_{\boxed{2}}$$\rangle$

$\stackrel{\pi_8}{\Rightarrow}$ $\langle$Aozhou shi X, Australia is X$\rangle$

$\stackrel{\pi_5}{\Rightarrow}$ $\langle$Aozhou shi X zhiyi, Australia is one of X$\rangle$

$\stackrel{\pi_4}{\Rightarrow}$ $\langle$Aozhou shi $X_{\boxed{1}}$ de $X_{\boxed{2}}$ zhiyi,
Australia is one of the $X_{\boxed{2}}$ that $X_{\boxed{1}}$$\rangle$

$\stackrel{\pi_3}{\Rightarrow}$ $\langle$Aozhou shi yu $X_{\boxed{1}}$ you $X_{\boxed{0}}$ de $X_{\boxed{2}}$ zhiyi,
Australia is one of the $X_{\boxed{2}}$ that have $X_{\boxed{0}}$ with $X_{\boxed{1}}$$\rangle$

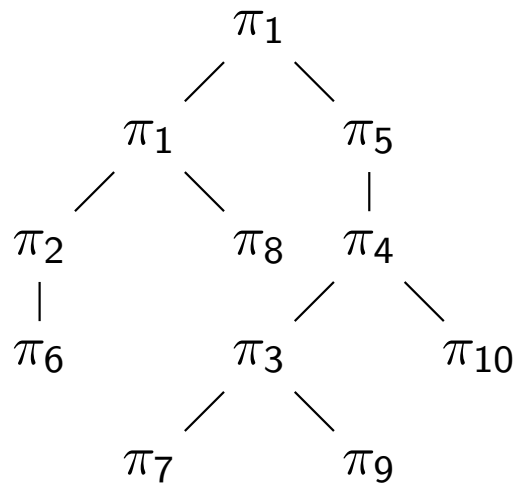
$\stackrel{\pi_7}{\Rightarrow}$ $\langle$Aozhou shi yu Beihan you $X_{\boxed{0}}$ de $X_{\boxed{2}}$ zhiyi,
Australia is one of the $X_{\boxed{2}}$ that have $X_{\boxed{0}}$ with North Korea$\rangle$

$\stackrel{\pi_9}{\Rightarrow}$ $\langle$Aozhou shi yu Beihan you bangjiao de $X_{\boxed{2}}$ zhiyi,
Australia is one of the $X_{\boxed{2}}$ that have diplomatic relations with North Korea$\rangle$

$\stackrel{\pi_{10}}{\Rightarrow}$ $\langle$Aozhou shi yu Beihan you bangjiao de shaoshu guojia zhiyi,
Australia is one of the few countries that have diplomatic relations with N. K.$\rangle$

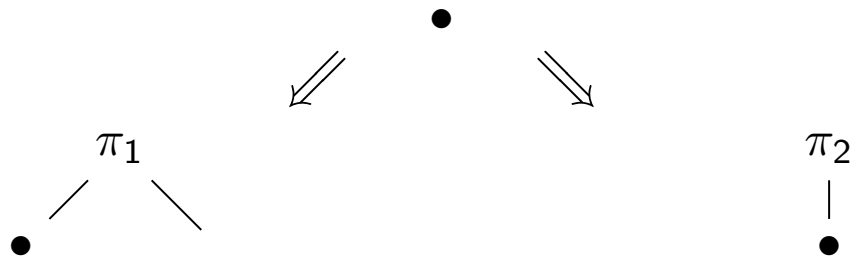
Inducing a Model

Derivation Tree



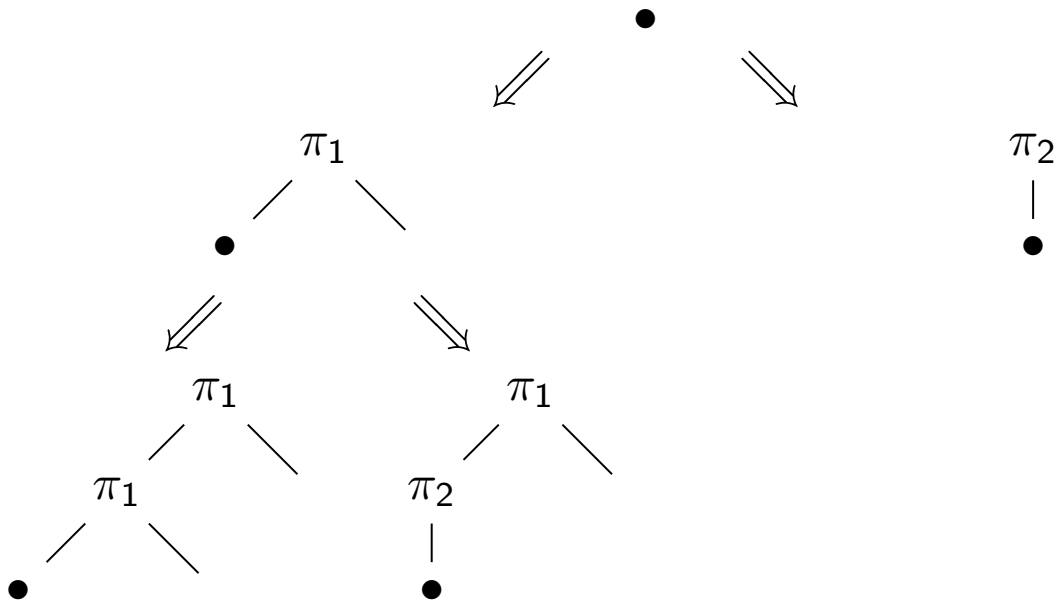
Inducing a Model

Decision Tree



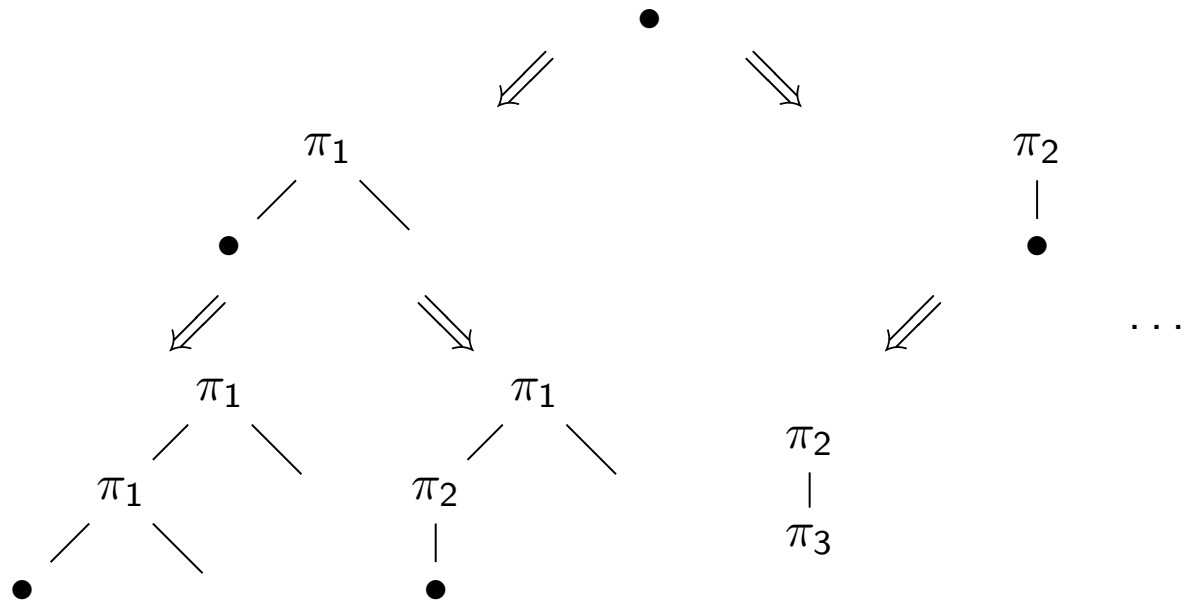
Inducing a Model

Decision Tree



Inducing a Model

Decision Tree



Inducing a Model

Assignments

$$\left.\begin{array}{r}\text{grammar } G \\ \text{assignment } p'\end{array}\right\} \text{“distribution”} \;\; p'(e, d, f)$$

$$\text{model } \mathcal{M}_G = \{p \in \mathcal{M}(E, F) \mid \exists p' \colon p(e, f) = \textstyle\sum_d p'(e, d, f)\}$$

Training: Rule Extraction

Word-aligned Training Pair

⟨30 duonianlai de youhao hezuo, friendly cooperation over the last 30 years⟩

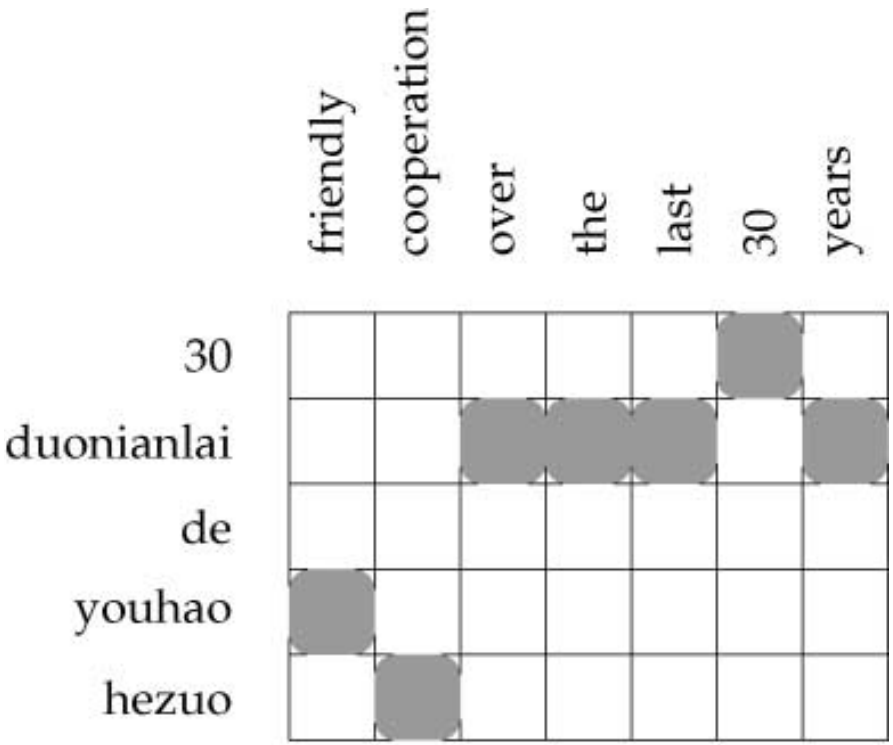


Figure 2 of (Chiang 2007)

Training: Rule Extraction

Initial Rules

$X \rightarrow \langle$30 duonianlai de youhao hezuo, friendly cooperation over the last 30 years$\rangle$

$X \rightarrow \langle$duonianlai, over the last 30 years$\rangle$

$X \rightarrow \langle$youhao hezuo, friendly cooperation$\rangle$

$X \rightarrow \langle 30, 30 \rangle$

$X \rightarrow \langle$youhao, friendly$\rangle$ $\quad$ $X \rightarrow \langle$hezuo, cooperation$\rangle$

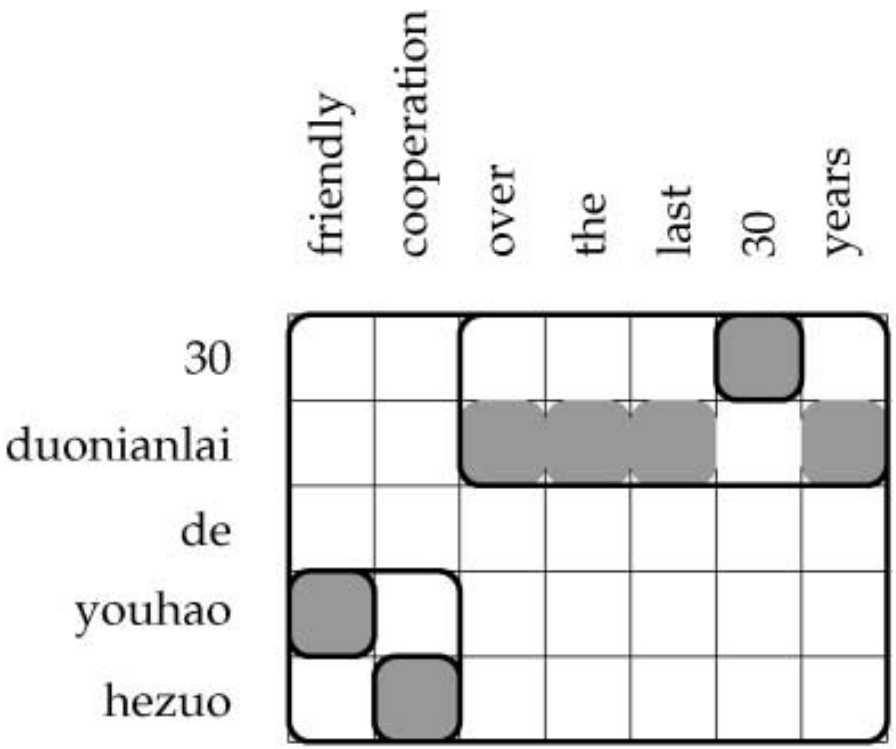


Figure 2 of (Chiang 2007)

Training: Rule Extraction

Example Rule

$X \rightarrow \langle X_{[1]}$ duonianlai de $X_{[2]}$, $X_{[2]}$ over the last $X_{[1]}$ years$\rangle$

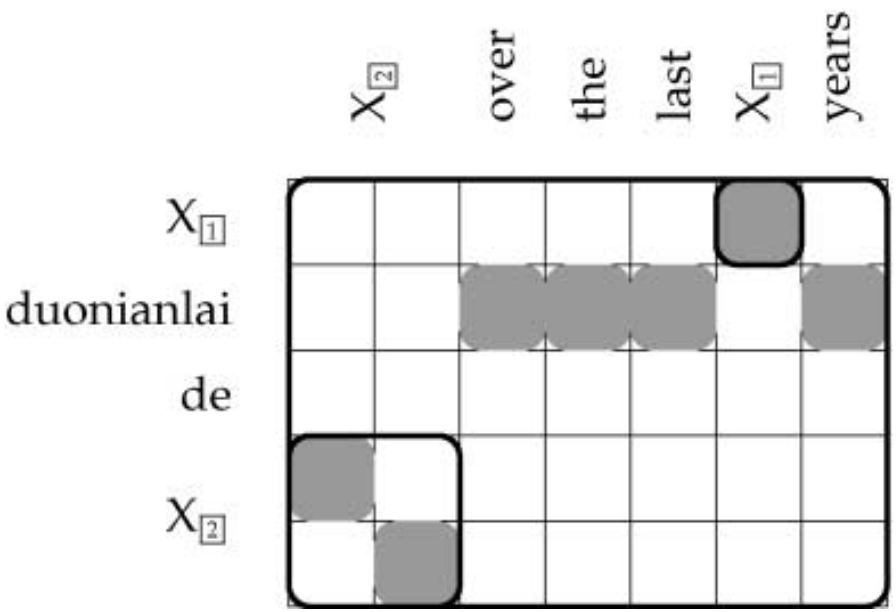


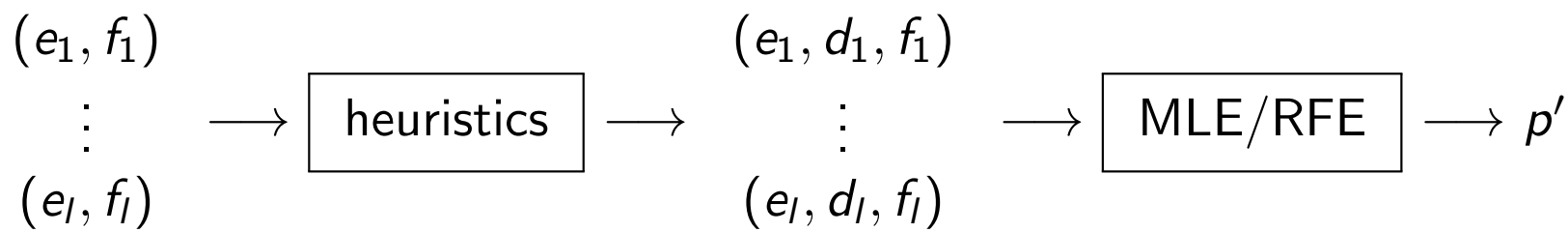
Figure 2 of (Chiang 2007)

Training: Maximum-Likelihood Estimation

- training data: sequence $\mathcal{D} = (e_1, f_1), \ldots, (e_l, f_l)$
- model: $\mathcal{M}_G = \{p \in \mathcal{M}(E, F) \mid \exists p' : p(e, f) = \sum_d p'(e, d, f)\}$
- likelihood: $P(\mathcal{D} \mid p') = \prod_j \sum_d p'(e_j, d, f_j)$
- use EM algorithm

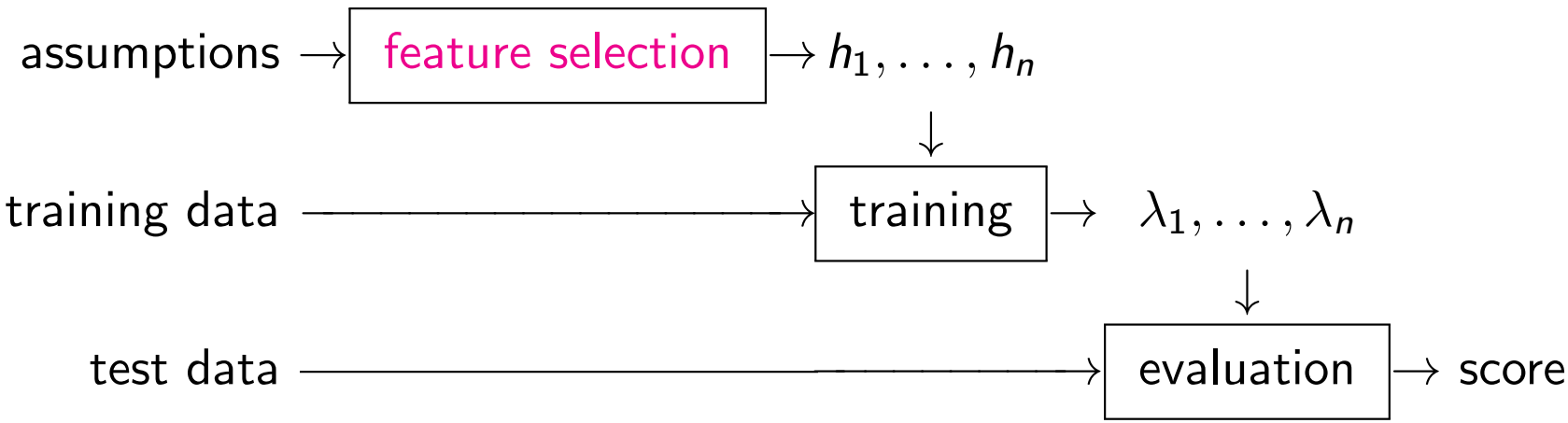
Training: Reality

Chiang (2007), DeNeefe and Knight (2009):



Outline

Building an SMT System with Log-linear Models



Feature Selection

Hidden variable: derivation of (e, f) in G

Feature Selection

Hidden variable: derivation of (e, f) in G

- $\mathcal{M} \subseteq \mathcal{M}(E \mid F)$
- $h_1(e, f) = \log \sum_d p_G(e, d, f)$
- $h_2(e, f) = \log p(e)$
- $h_3(e, f) = |e|$

Feature Selection

Hidden variable: derivation of (e, f) in G

- $\mathcal{M} \subseteq \mathcal{M}(E \mid F)$
- $h_1(e, f) = \log \sum_d p_G(e, d, f)$
- $h_2(e, f) = \log p(e)$
- $h_3(e, f) = |e|$
- $h_4(e, f) = \log \sum_a p(f, a \mid e)$
  alignment-based model

Feature Selection

Hidden variable: derivation of (e, f) in G

- $\mathcal{M} \subseteq \mathcal{M}(E \mid F)$
- $h_1(e, f) = \log \sum_d p_G(e, d, f)$
- $h_2(e, f) = \log p(e)$
- $h_3(e, f) = |e|$
- $h_4(e, f) = \log \sum_a p(f, a \mid e)$

  alignment-based model

Chiang (2007):

- $\mathcal{M} \subseteq \mathcal{M}(E, D \mid F)$
- $h_1(e, d, f) = \log p_G(e, d \mid f)$
- $h_2(e, d, f) = \log p_G(f, d \mid e)$
- $h_3(e, d, f) = \log p(e)$
- $h_4(e, d, f) = |e|$

Feature Selection

Hidden variable: derivation of (e, f) in G

- $\mathcal{M} \subseteq \mathcal{M}(E \mid F)$
- $h_1(e, f) = \log \sum_d p_G(e, d, f)$
- $h_2(e, f) = \log p(e)$
- $h_3(e, f) = |e|$
- $h_4(e, f) = \log \sum_a p(f, a \mid e)$

  alignment-based model

Chiang (2007):

- $\mathcal{M} \subseteq \mathcal{M}(E, D \mid F)$
- $h_1(e, d, f) = \log p_G(e, d \mid f)$
- $h_2(e, d, f) = \log p_G(f, d \mid e)$
- $h_3(e, d, f) = \log p(e)$
- $h_4(e, d, f) = |e|$
- $h_5(e, d, f) =$ lexical weights

Feature Selection

Hidden variable: derivation of (e, f) in G

- $\mathcal{M} \subseteq \mathcal{M}(E \mid F)$
- $h_1(e, f) = \log \sum_d p_G(e, d, f)$
- $h_2(e, f) = \log p(e)$
- $h_3(e, f) = |e|$
- $h_4(e, f) = \log \sum_a p(f, a \mid e)$
  alignment-based model

Chiang (2007):

- $\mathcal{M} \subseteq \mathcal{M}(E, D \mid F)$
- $h_1(e, d, f) = \log p_G(e, d \mid f)$
- $h_2(e, d, f) = \log p_G(f, d \mid e)$
- $h_3(e, d, f) = \log p(e)$
- $h_4(e, d, f) = |e|$
- $h_5(e, d, f) =$ lexical weights
- $h_6(e, d, f) = \#_{\text{ex}} d$
- $h_7(e, d, f) = \#_{\text{gl}} d$

Outline

Decoding

$$\hat{e} = \text{argmax}_e \; p(e \mid f)$$

Decoding

$$\begin{aligned}\hat{e} &= \text{argmax}_e \; p(e \mid f) \\ &= \text{argmax}_e \sum_d p(e, d \mid f)\end{aligned}$$

Decoding

$$
\begin{aligned}
\hat{e} &= \text{argmax}_e \; p(e \mid f) \\
&= \text{argmax}_e \sum_d p(e, d \mid f) \\
&= \text{argmax}_e \sum_d \frac{\exp \sum_i \lambda_i \cdot h_i(e, d, f)}{\sum_{e', d'} \exp \sum_i \lambda_i \cdot h_i(e', d', f)}
\end{aligned}
$$

Decoding

$$
\begin{aligned}
\hat{e} &= \text{argmax}_e \; p(e \mid f) \\
&= \text{argmax}_e \sum_d p(e, d \mid f) \\
&= \text{argmax}_e \sum_d \frac{\exp \sum_i \lambda_i \cdot h_i(e, d, f)}{\sum_{e', d'} \exp \sum_i \lambda_i \cdot h_i(e', d', f)} \\
&= \text{argmax}_e \sum_d \exp \sum_i \lambda_i \cdot h_i(e, d, f)
\end{aligned}
$$

Decoding

$$\begin{aligned}
\hat{e} &= \text{argmax}_e \; p(e \mid f) \\
&= \text{argmax}_e \sum_d p(e, d \mid f) \\
&= \text{argmax}_e \sum_d \frac{\exp \sum_i \lambda_i \cdot h_i(e, d, f)}{\sum_{e', d'} \exp \sum_i \lambda_i \cdot h_i(e', d', f)} \\
&= \text{argmax}_e \sum_d \exp \sum_i \lambda_i \cdot h_i(e, d, f)
\end{aligned}$$

NP-hard! (Casacuberta and Higuera 2000; Li, Eisner, and Khudanpur 2009)

Decoding

$$\hat{e} = \text{argmax}_e \; p(e \mid f)$$

$$= \text{argmax}_e \sum_d p(e, d \mid f)$$

$$= \text{argmax}_e \sum_d \frac{\exp \sum_i \lambda_i \cdot h_i(e, d, f)}{\sum_{e', d'} \exp \sum_i \lambda_i \cdot h_i(e', d', f)}$$

$$= \text{argmax}_e \sum_d \exp \sum_i \lambda_i \cdot h_i(e, d, f)$$

NP-hard! (Casacuberta and Higuera 2000; Li, Eisner, and Khudanpur 2009)

$$\approx \text{argmax}_e \sum_{d \in D_{\text{fin}}} \exp \sum_i \lambda_i \cdot h_i(e, d, f)$$

Decoding

$$\hat{e} = \text{argmax}_e \; p(e \mid f)$$

$$= \text{argmax}_e \sum_d p(e, d \mid f)$$

$$= \text{argmax}_e \sum_d \frac{\exp \sum_i \lambda_i \cdot h_i(e, d, f)}{\sum_{e', d'} \exp \sum_i \lambda_i \cdot h_i(e', d', f)}$$

$$= \text{argmax}_e \sum_d \exp \sum_i \lambda_i \cdot h_i(e, d, f)$$

NP-hard! (Casacuberta and Higuera 2000; Li, Eisner, and Khudanpur 2009)

$$\approx \text{argmax}_e \sum_{d \in D_{\text{fin}}} \exp \sum_i \lambda_i \cdot h_i(e, d, f)$$

where $D_{\text{fin}} = \text{argmax}_d^{(n)} \exp \sum_i \lambda_i \cdot h_i(\text{e}(d), d, f)$

Computing D_{fin}

$$D_{\text{fin}} = \text{argmax}_d^{(n)} \exp \sum_i \lambda_i \cdot h_i(\text{e}(d), d, f)$$

Computing D_{fin}

$$D_{\text{fin}} = \text{argmax}_d^{(n)} \exp \sum_i \lambda_i \cdot h_i(\text{e}(d), d, f)$$

$$\max\{\ 0.9,\ 0.99,\ 0.999,\ \ldots\ \}\ ?$$

Computing D_{fin}

$$D_{\text{fin}} = \text{argmax}_d^{(n)} \exp \sum_i \lambda_i \cdot h_i(\text{e}(d), d, f)$$

$$G' \rightarrow \boxed{\text{n-best}} \rightarrow \text{argmax}_d^{(n)}\, p_{G'}(d)$$

n-best $\approx O(|R'| + |N'| \cdot n \cdot \log n)$ (Huang and Chiang 2005)

Computing D_{fin}

$$D_{\text{fin}} = \text{argmax}_d^{(n)} \exp \sum_i \lambda_i \cdot h_i(\text{e}(d), d, f)$$

$$\begin{array}{r} G \to \\ f \to \end{array} \boxed{\begin{array}{c}\text{input} \\ \text{product}\end{array}} \to G' \to \boxed{\text{n-best}} \to \begin{array}{l} \text{argmax}_d^{(n)}\, p_{G'}(d) \\ = \text{argmax}_{d:\, \text{f}(d)=f}^{(n)}\, p_G(d) \end{array}$$

$$\begin{aligned} \text{n-best} &\approx O(|R'| + |N'| \cdot n \cdot \log n) \text{ (Huang and Chiang 2005)} \\ \text{input product} &\approx O(|R| \cdot |f|^c) \end{aligned}$$

Computing D_{fin}

$$D_{\text{fin}} = \text{argmax}_d^{(n)} \exp \sum_i \lambda_i \cdot h_i(\text{e}(d), d, f)$$

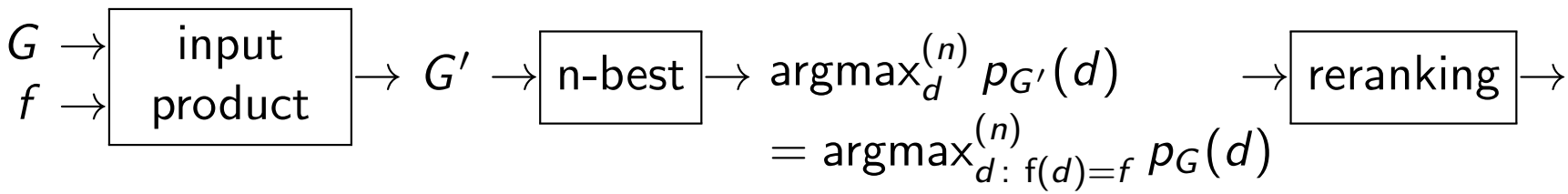
$$\begin{array}{l} G \to \\ f \to \end{array} \boxed{\begin{array}{c}\text{input}\\ \text{product}\end{array}} \to G' \to \boxed{\text{n-best}} \to \begin{array}{l} \text{argmax}_d^{(n)}\, p_{G'}(d) \\ \quad = \text{argmax}_{d:\, \text{f}(d)=f}^{(n)}\, p_G(d) \end{array}$$

$$\begin{aligned} \text{n-best} &\approx O(|R'| + |N'| \cdot n \cdot \log n) \text{ (Huang and Chiang 2005)} \\ \text{input product} &\approx O(|R| \cdot |f|^c) \end{aligned}$$

$$\rightsquigarrow \text{Cube Pruning (Chiang 2007)}$$

Computing D_{fin}

$$D_{\text{fin}} = \text{argmax}_d^{(n)} \exp \sum_i \lambda_i \cdot h_i(\text{e}(d), d, f)$$

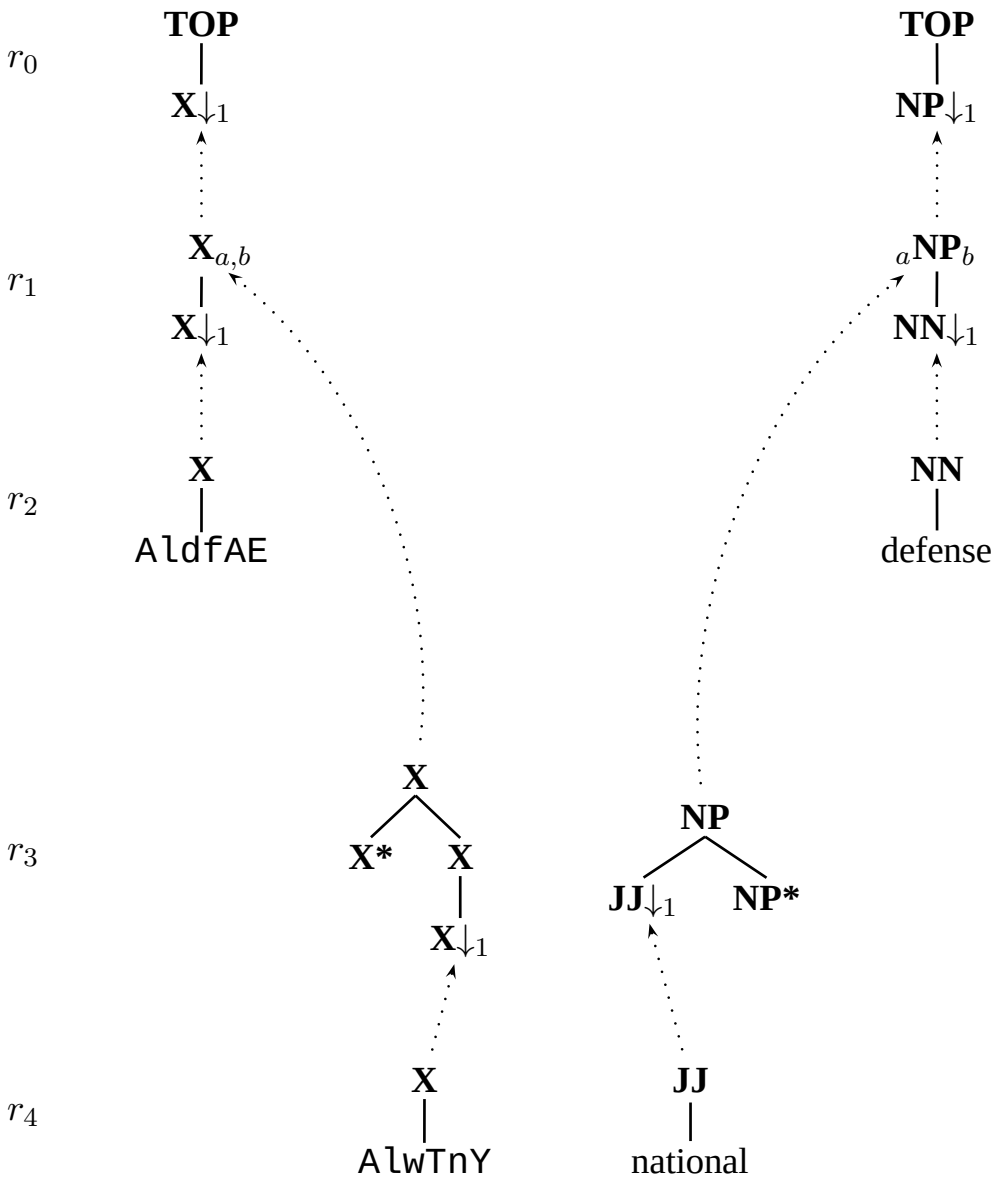


$$\text{n-best} \approx O(|R'| + |N'| \cdot n \cdot \log n) \text{ (Huang and Chiang 2005)}$$
$$\text{input product} \approx O(|R| \cdot |f|^c)$$

$$\rightsquigarrow \text{Cube Pruning (Chiang 2007)}$$

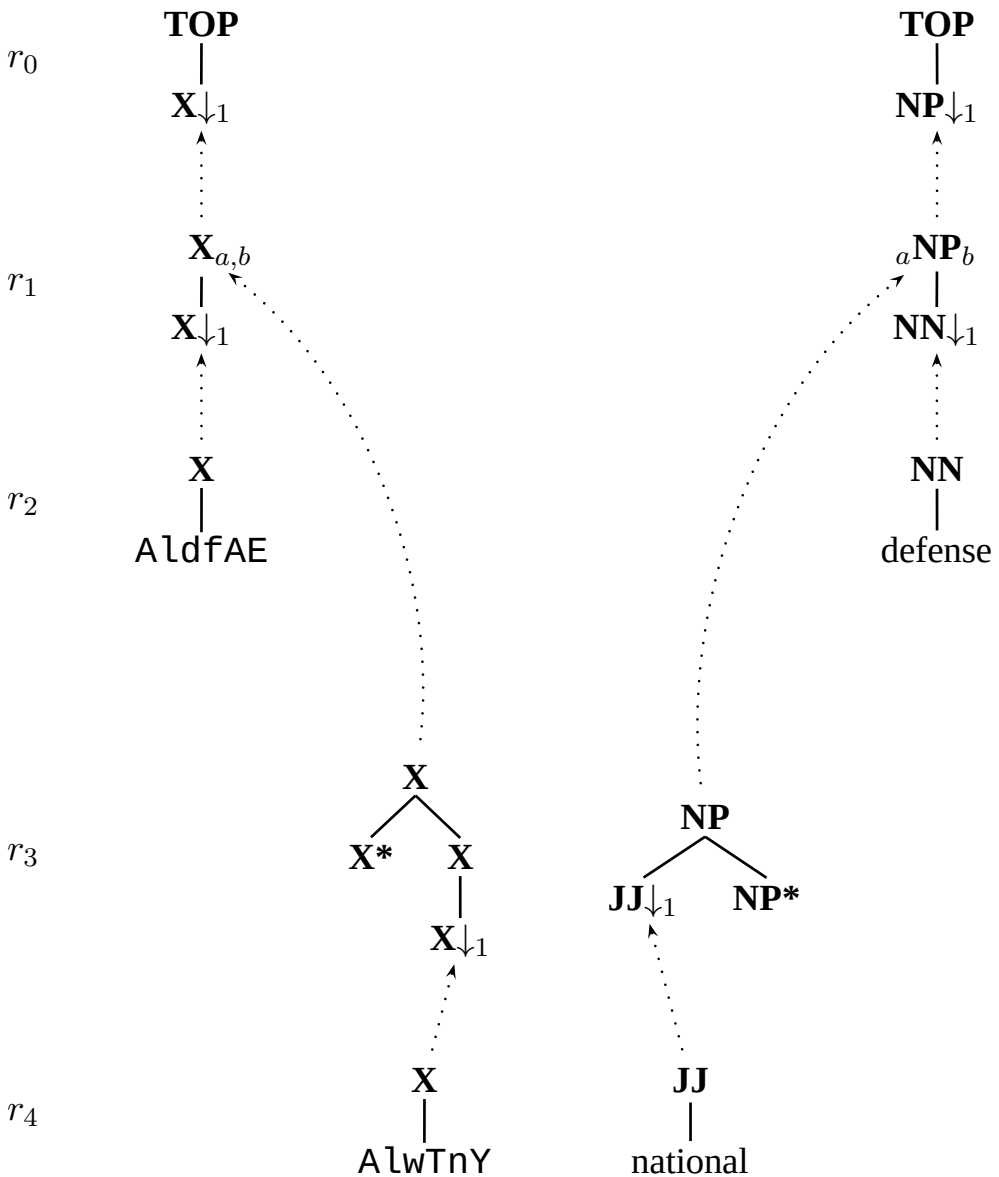
Outline

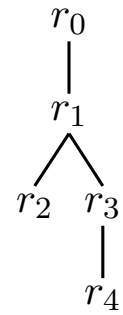
Example: Synchronous Tree-Insertion Grammar



(DeNeefe 2009)

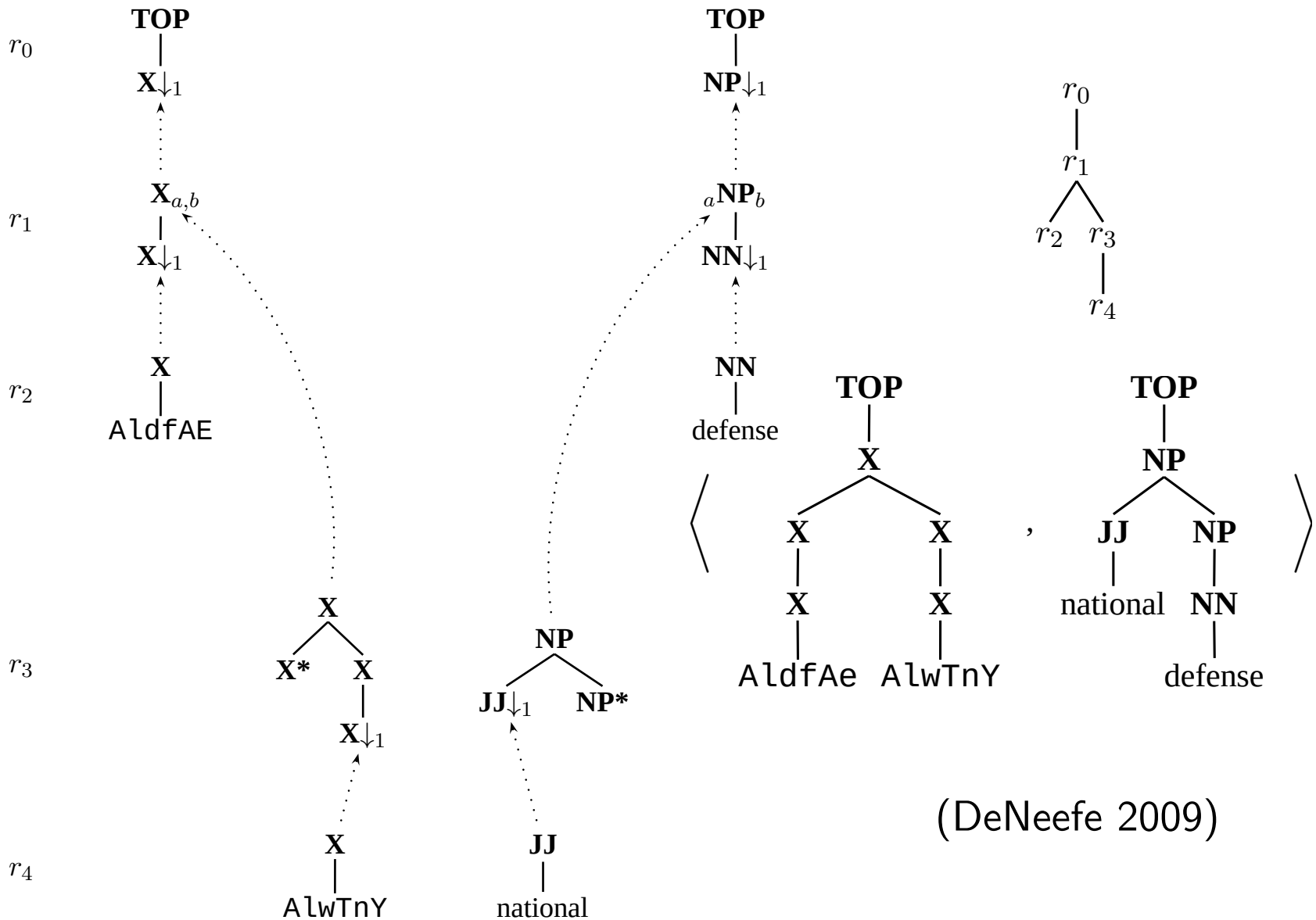
Example: Synchronous Tree-Insertion Grammar





(DeNeefe 2009)

Example: Synchronous Tree-Insertion Grammar



(DeNeefe 2009)

Overview

- Extended Top-down Tree-to-String Transducers (Galley et al. 2004; Huang, Knight, and Joshi 2006; Graehl, Knight, and May 2008)
- Extended Multi-Bottom-up Tree Transducers (Engelfriet, Lilin, and Maletti 2008)
- Synchronous Context-Free Grammars (Chiang 2007)
- Synchronous Tree-Substitution Grammars (Eisner 2003)
- Synchronous Tree-Sequence-Substitution Grammars (Zhang et al. 2008)
- Synchronous Tree-Insertion Grammars (Nesson, Shieber, and Rush 2006; DeNeefe and Knight 2009)

Outline

Key Aspects

- Translation Model: Synchronous Tree-Adjoining Grammars
  - formulation with states and explicit sites
  - formal properties: closure under input/output product
  - algorithms for training, decoding

Key Aspects

- Translation Model: Synchronous Tree-Adjoining Grammars
  - formulation with states and explicit sites
  - formal properties: closure under input/output product
  - algorithms for training, decoding
- Auxiliary Algorithms for Decoding
  - n-best algorithm
  - determinization of weighted tree automata

Outline

Publications so far

- MB, Jonathan May, and Heiko Vogler (2010).
  "Determinization of Weighted Tree Automata using Factorizations".
  In: Journal of Automata, Languages and Combinatorics. (accepted for publication)
- MB, Daniel Geisler, Torsten Stüber, and Heiko Vogler (2010).
  "n-Best Parsing Revisited".
  In: Proceedings of the 2010 Workshop on Applications of Tree Automata in Natural Language Processing, ACL 2010. Uppsala, Sweden, 16 July 2010, pp. 46–54.
- MB, Mark-Jan Nederhof, and Heiko Vogler (2011).
  "Tree Parsing with Synchronous Tree-Adjoining Grammars".
  In: Proceedings of IWPT. (accepted for publication)