As Easy As Vanda, Two, Three: Components for Machine Translation Based on Formal Grammars

Matthias Büchse

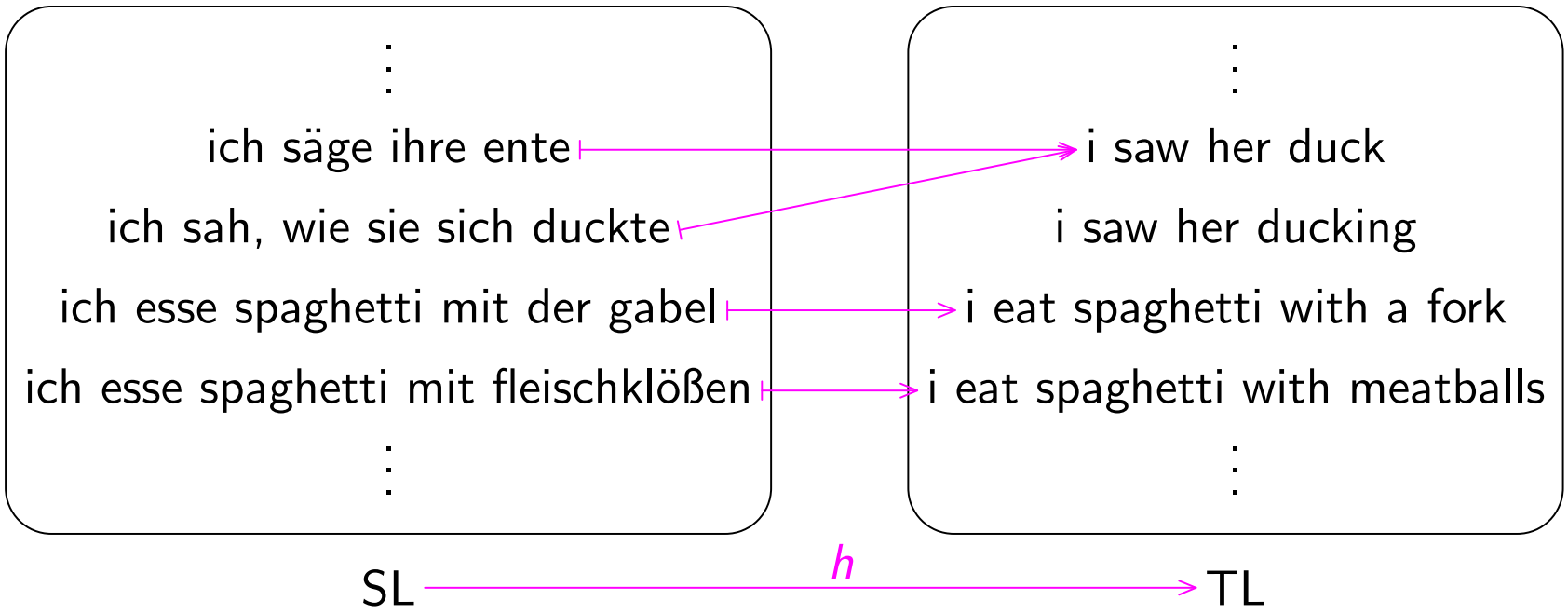
TECHNISCHE UNIVERSITÄT DRESDEN

Theorietag, Prague, 2012-10-04

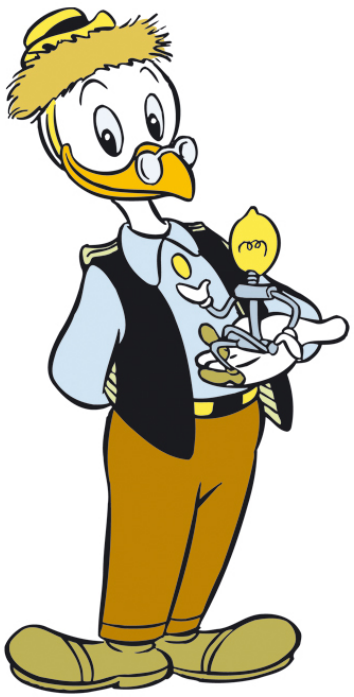
Outline

Outline

Goal


ich säge ihre ente
ich sah, wie sie sich duckte
ich esse spaghetti mit der gabel
ich esse spaghetti mit fleischklößen
i saw her duck
i saw her ducking
i eat spaghetti with a fork
i eat spaghetti with meatballs
SL
TL
h

Modelling and Algorithmisation

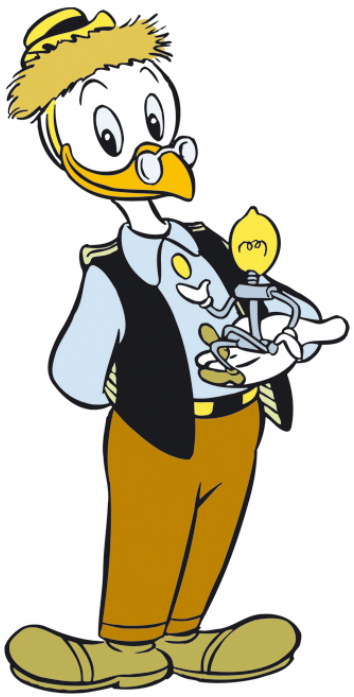


$S \to \langle \text{ich säge } X, \text{I saw } X \rangle$
$X \to \langle \text{ihre Ente}, \text{her duck} \rangle$

$S \Rightarrow \cdots \Rightarrow \langle \text{Ich säge ihre Ente}, \text{I saw her duck} \rangle$

modelling — select $\mathcal{H} \subseteq \{h \mid h\colon \text{SL} \to \text{TL}\}$
e.g., via synchronous grammar

Modelling and Algorithmisation



$S \to \langle \text{ich säge } X, \text{I saw } X \rangle$
$X \to \langle \text{ihre Ente}, \text{her duck} \rangle$

$S \Rightarrow \cdots \Rightarrow \langle \text{Ich säge ihre Ente}, \text{I saw her duck} \rangle$

modelling select $\mathcal{H} \subseteq \{h \mid h\colon \mathsf{SL} \to \mathsf{TL}\}$
e.g., via synchronous grammar

```
program Decoder;
...
begin
  ...
end.
```

algorithmise e.g., decoder:
given $h \in \mathcal{H}$, $s \in \mathsf{SL}$
compute $h(s)$

Training

sentence-aligned bilingual data: parallel corpus

001 Resumption of the session
002 I declare resumed the session of the European Parliament adjourned on Friday 17 December 1999 , [. . .] .

001 Wiederaufnahme der Sitzungsperiode
002 Ich erkläre die am Freitag , dem 17. Dezember unterbrochene Sitzungsperiode des Europäischen Parlaments für wiederaufgenommen , [. . .] .

EuroParl corpus, 11 languages, 1.5M sentences each

Training

sentence-aligned bilingual data: parallel corpus

↓

apply heuristic, statistical methods

Training

sentence-aligned bilingual data: parallel corpus

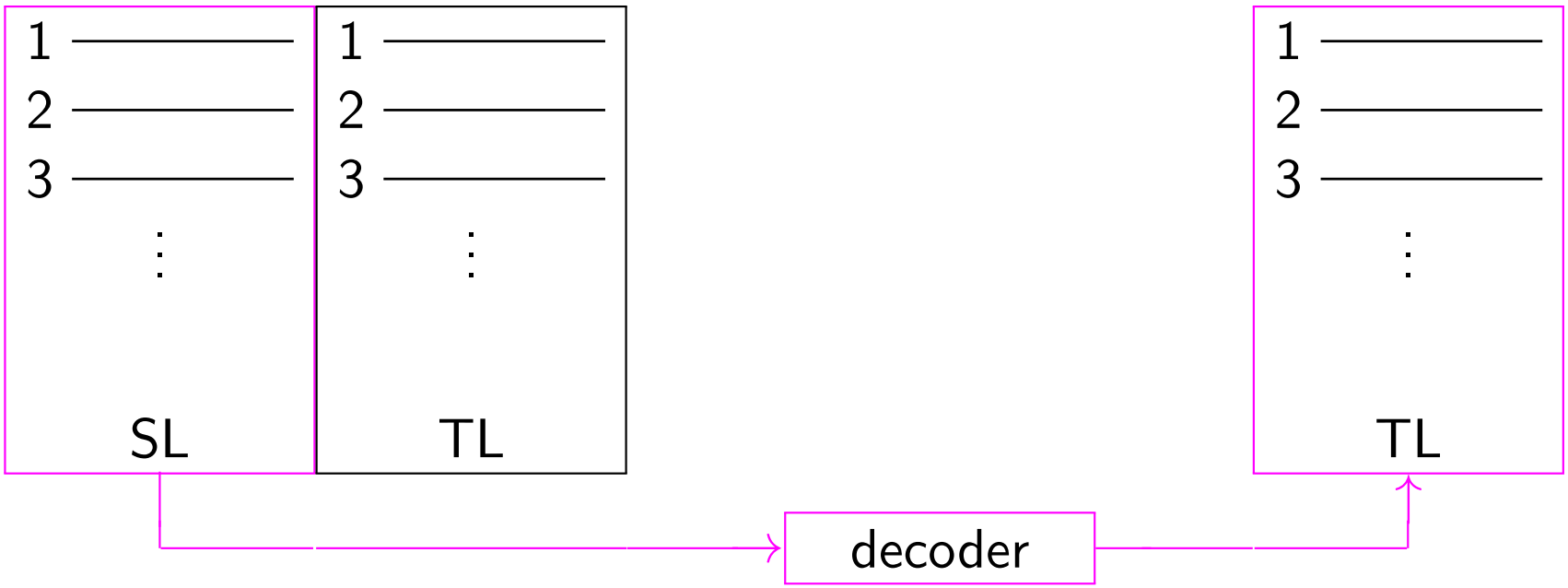
↓

apply heuristic, statistical methods

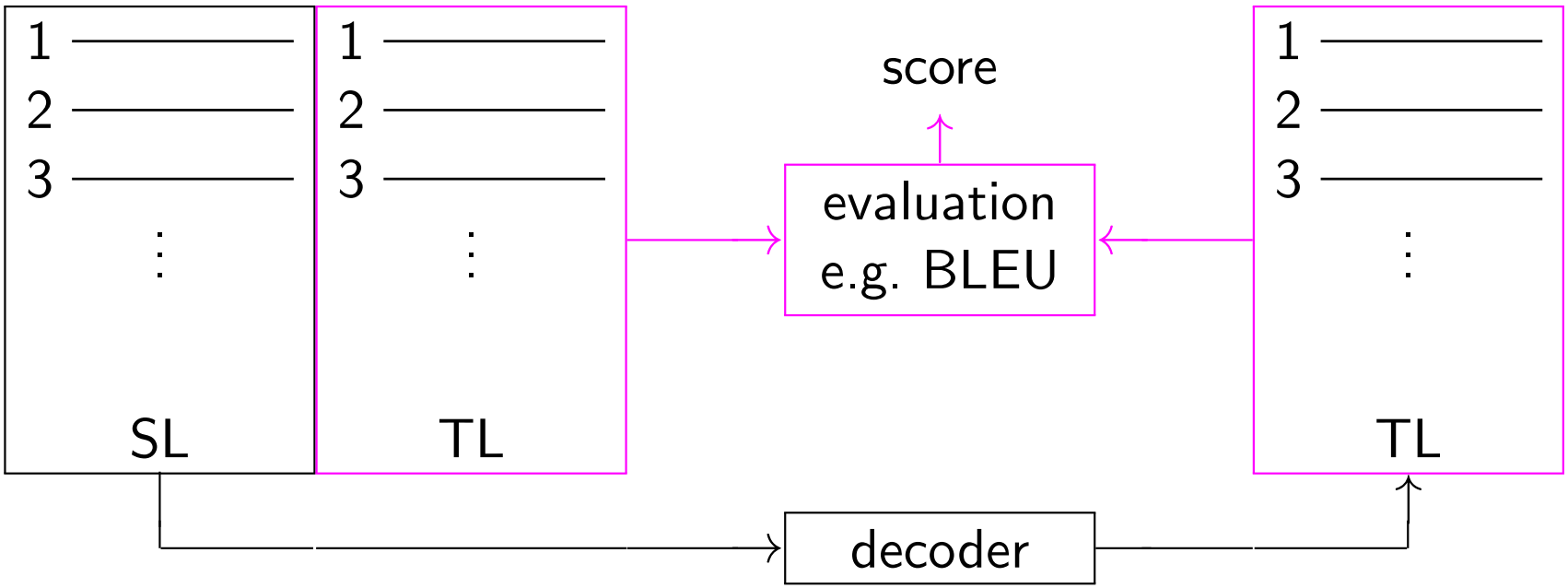
↓

grammar rules, weights

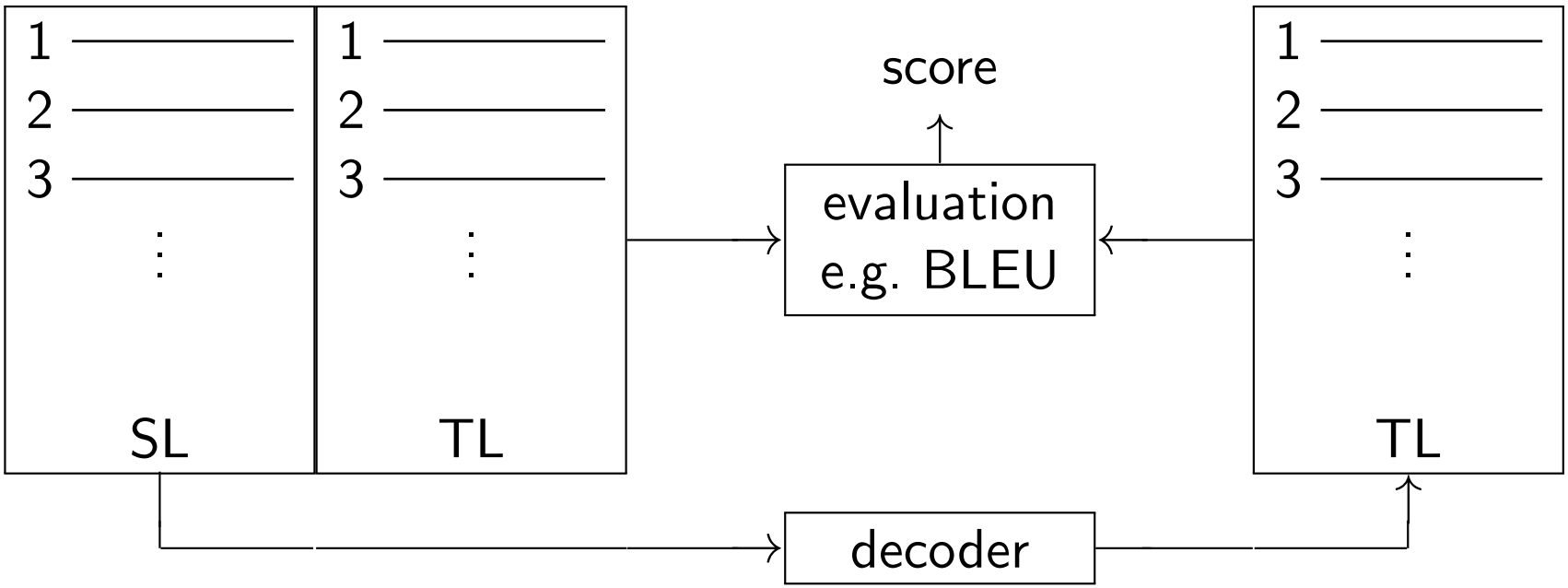
Evaluation



Evaluation



Evaluation



if score $>$ oldscore **then** publish **else** perish

Outline

Synchronous Context-Free Grammar

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

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

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

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

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

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

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

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

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

Derivation

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

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

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

Derivation

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

Derivation

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

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$

$\overset{\pi_6}{\Rightarrow}$ $\langle$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 . . .$\rangle$

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\ \ldots} \rangle \\
\overset{\pi_{10}}{\Rightarrow}\ & \langle \mathrm{Aozhou\ shi\ yu\ Beihan\ you\ bangjiao\ de\ shaoshu\ guojia\ zhiyi}, \\
& \quad \mathrm{Australia\ is\ one\ of\ the\ few\ countries\ that\ have\ \ldots} \rangle
\end{aligned}
$$

Derivation

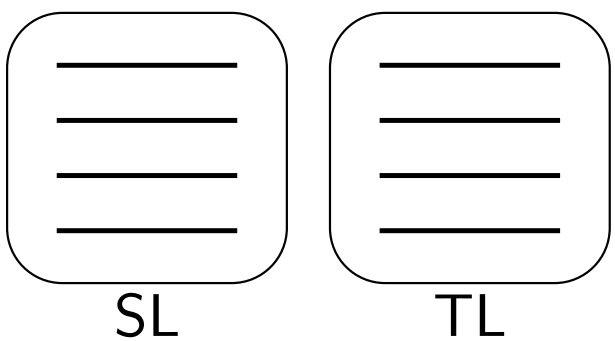
$$
\begin{aligned}
& \langle \mathrm{S},\ \mathrm{S} \rangle \\
\overset{\pi_1}{\Rightarrow}\ & \langle \mathrm{S\ X},\ \mathrm{S\ X} \rangle \\
\overset{\pi_1}{\Rightarrow}\ & \langle \mathrm{S\ X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}},\ \mathrm{S\ X}_{\boxed{1}}\ \mathrm{X}_{\boxed{2}} \rangle \\
\overset{\pi_2}{\Rightarrow}\ & \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\ \ldots} \rangle \\
\overset{\pi_{10}}{\Rightarrow}\ & \langle \mathrm{Aozhou\ shi\ yu\ Beihan\ you\ bangjiao\ de\ shaoshu\ guojia\ zhiyi}, \\
& \quad \mathrm{Australia\ is\ one\ of\ the\ few\ countries\ that\ have\ \ldots} \rangle
\end{aligned}
$$

Model

$$\mathcal{H} = \{h_{G,\Phi,\theta} \mid G \in \mathcal{G}, \Phi \colon D(G) \to \mathbb{R}^m, \theta \in \mathbb{R}^m\}$$

where

$$h_{G,\Phi,\theta} \colon s \mapsto$$

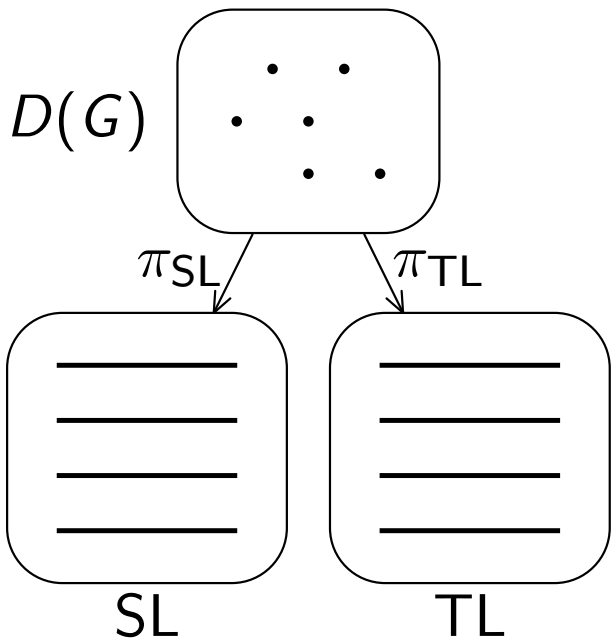


Model

$$\mathcal{H} = \{h_{G,\Phi,\theta} \mid G \in \mathcal{G}, \Phi \colon D(G) \to \mathbb{R}^m, \theta \in \mathbb{R}^m\}$$

where

$$h_{G,\Phi,\theta} \colon s \mapsto$$

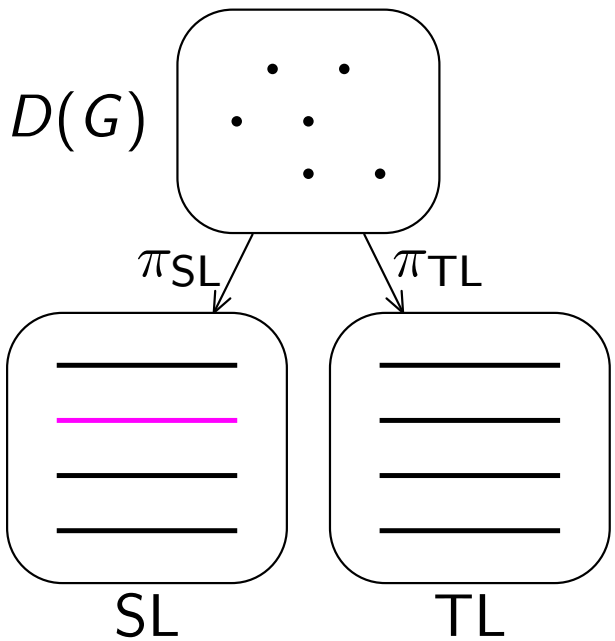


Model

$$\mathcal{H} = \{h_{G,\Phi,\theta} \mid G \in \mathcal{G}, \Phi\colon D(G) \to \mathbb{R}^m, \theta \in \mathbb{R}^m\}$$

where

$$h_{G,\Phi,\theta}\colon s \mapsto$$

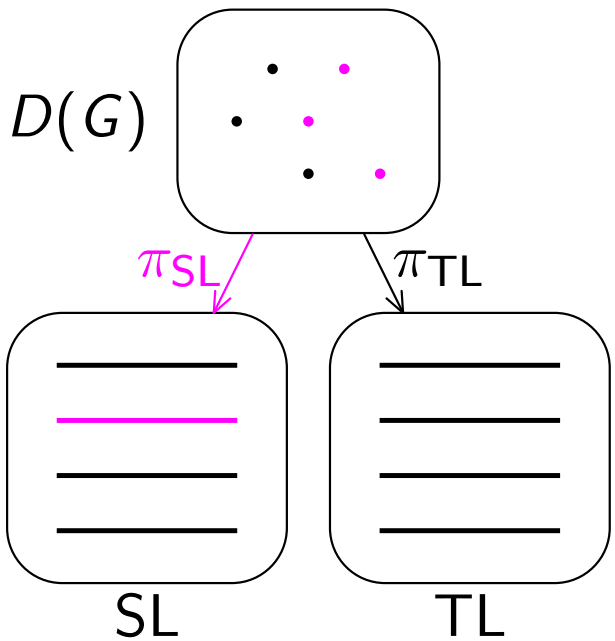


Model

$$\mathcal{H} = \{h_{G,\Phi,\theta} \mid G \in \mathcal{G}, \Phi\colon D(G) \to \mathbb{R}^m, \theta \in \mathbb{R}^m\}$$

where

$$h_{G,\Phi,\theta}\colon s \mapsto$$

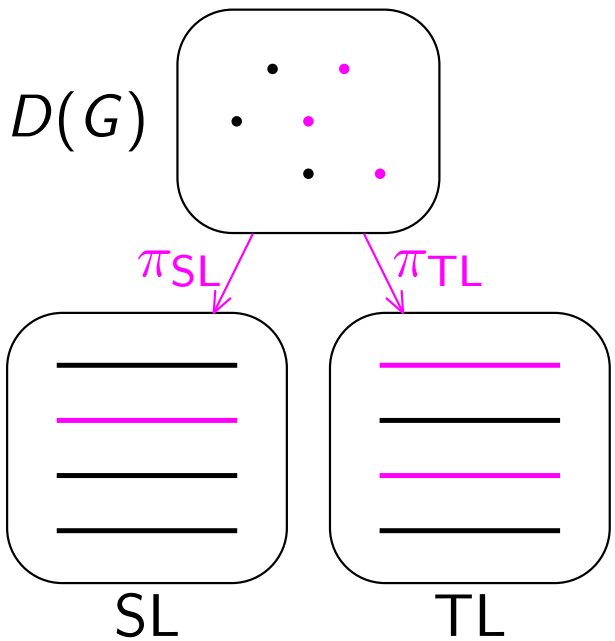


Model

$$\mathcal{H} = \{h_{G,\Phi,\theta} \mid G \in \mathcal{G}, \Phi\colon D(G) \to \mathbb{R}^m, \theta \in \mathbb{R}^m\}$$

where

$$h_{G,\Phi,\theta}\colon s \mapsto \pi_{\mathsf{TL}}\Big(\qquad \Big)$$

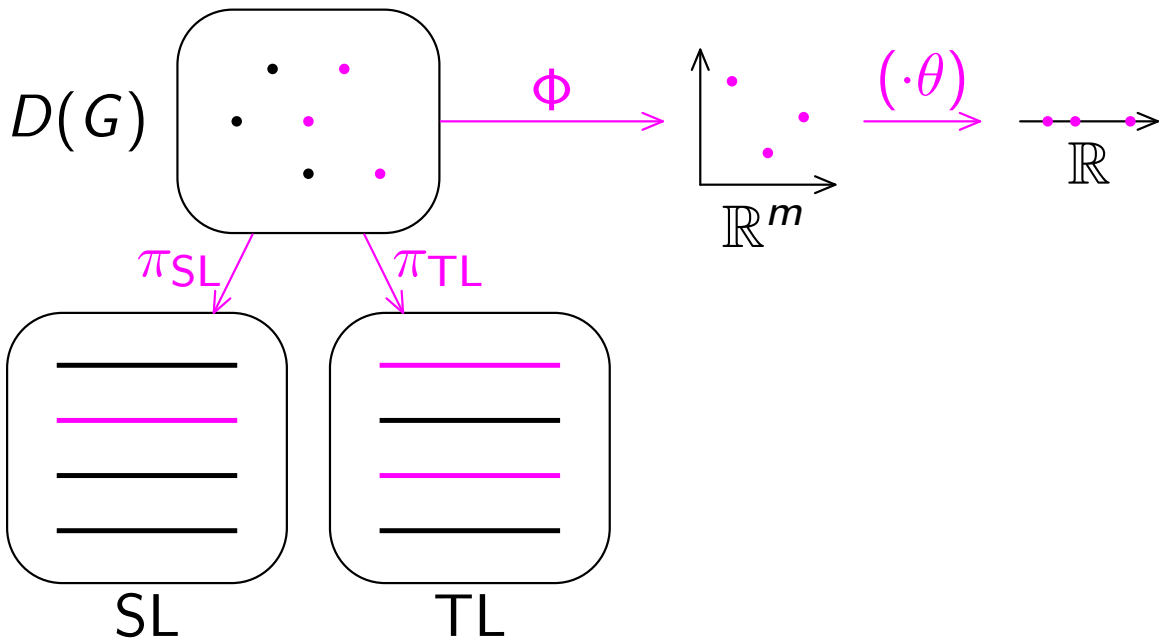


Model

$$\mathcal{H} = \{h_{G,\Phi,\theta} \mid G \in \mathcal{G}, \Phi\colon D(G) \to \mathbb{R}^m, \theta \in \mathbb{R}^m\}$$

where

$$h_{G,\Phi,\theta}\colon s \mapsto \pi_{\mathsf{TL}}\Big(\qquad\qquad [\![G]\!](d) + \Phi(d)\cdot\theta \Big)$$

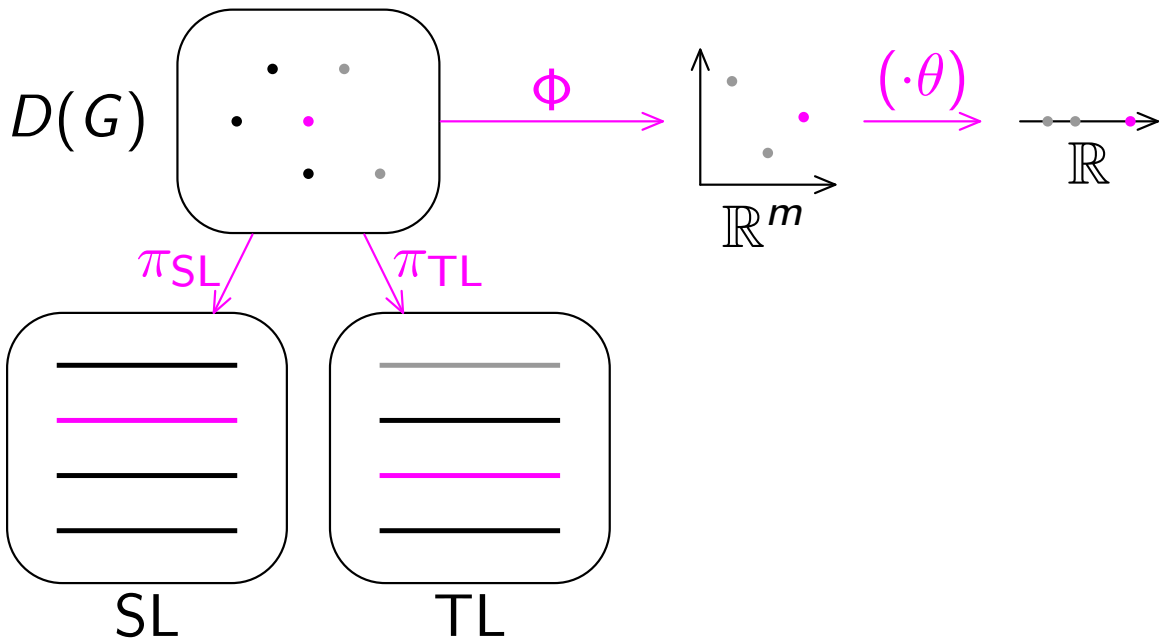


Model

$$\mathcal{H} = \{h_{G,\Phi,\theta} \mid G \in \mathcal{G}, \Phi\colon D(G) \to \mathbb{R}^m, \theta \in \mathbb{R}^m\}$$

where

$$h_{G,\Phi,\theta}\colon s \mapsto \pi_{\mathsf{TL}}\left(\mathsf{argmax}_{d\colon \pi_{\mathsf{SL}}(d)=s}\ [\![G]\!](d) + \Phi(d)\cdot\theta\right)$$



SCFG Implementations

name	language	BLEU	speed	reference
Hiero	Python	31.22	27.2 s/sent	Chiang et al. 2005
Joshua	Java	31.55	2.34 s/sent	Li et al. 2009
cdec	C++	31.47	0.37 s/sent	Dyer et al. 2010
Moses	C++	?	?	Koehn et al. 2007

SCFG Implementations

name	language	BLEU	speed	reference
Hiero	Python	31.22	27.2 s/sent	Chiang et al. 2005
Joshua	Java	31.55	2.34 s/sent	Li et al. 2009
cdec	C++	31.47	0.37 s/sent	Dyer et al. 2010
Moses	C++	?	?	Koehn et al. 2007

- BLEU score and speed trump adherance to specification

SCFG Implementations

name	language	BLEU	speed	reference
Hiero	Python	31.22	27.2 s/sent	Chiang et al. 2005
Joshua	Java	31.55	2.34 s/sent	Li et al. 2009
cdec	C++	31.47	0.37 s/sent	Dyer et al. 2010
Moses	C++	?	?	Koehn et al. 2007

- BLEU score and speed trump adherance to specification
- lacking conceptualization, vague presentation

SCFG Implementations

name	language	BLEU	speed	reference
Hiero	Python	31.22	27.2 s/sent	Chiang et al. 2005
Joshua	Java	31.55	2.34 s/sent	Li et al. 2009
cdec	C++	31.47	0.37 s/sent	Dyer et al. 2010
Moses	C++	?	?	Koehn et al. 2007

- BLEU score and speed trump adherance to specification
- lacking conceptualization, vague presentation
- concepts and code not reusable

SCFG Implementations

name	language	BLEU	speed	reference
Hiero	Python	31.22	27.2 s/sent	Chiang et al. 2005
Joshua	Java	31.55	2.34 s/sent	Li et al. 2009
cdec	C++	31.47	0.37 s/sent	Dyer et al. 2010
Moses	C++	?	?	Koehn et al. 2007

- BLEU score and speed trump adherance to specification
- lacking conceptualization, vague presentation
- concepts and code not reusable
- obstructs progress

SCFG Implementations

name	language	BLEU	speed	reference
Hiero	Python	31.22	27.2 s/sent	Chiang et al. 2005
Joshua	Java	31.55	2.34 s/sent	Li et al. 2009
cdec	C++	31.47	0.37 s/sent	Dyer et al. 2010
Moses	C++	?	?	Koehn et al. 2007

- BLEU score and speed trump adherance to specification
- lacking conceptualization, vague presentation
- concepts and code not reusable
- obstructs progress
- SCFGs, XTOPs, STSSG, STAG, MBOT, ...

Conventional Model and Decoder

POLISHED!

$$h_{G,\Phi,\theta} : s \mapsto \pi_{\mathsf{TL}}\left(\operatorname*{argmax}_{d\,:\,\pi_{\mathsf{SL}}(d)=s} [\![G]\!](d) + \Phi(d) \cdot \theta \right)$$

Conventional Model and Decoder POLISHED!

$$h_{G,\Phi,\theta} : s \mapsto \pi_{\mathsf{TL}}\Big(\operatorname*{argmax}_{d\,:\,\pi_{\mathsf{SL}}(d)=s} [\![G]\!](d) + \Phi(d)\cdot\theta\Big)$$

$$= \pi_{\mathsf{TL}}\big(\operatorname*{argmax}_{d} [\![(=s)]\!](\pi_{\mathsf{SL}}(d)) + [\![G]\!](d) + \Phi(d)\cdot\theta\big)$$

Conventional Model and Decoder POLISHED!

$$h_{G,\Phi,\theta} \colon s \mapsto \pi_{\mathsf{TL}}\Big(\operatorname*{argmax}_{d \colon \pi_{\mathsf{SL}}(d)=s} [\![G]\!](d) + \Phi(d) \cdot \theta\Big)$$

$$= \pi_{\mathsf{TL}}\big(\operatorname*{argmax}_{d} [\![(= s)]\!](\pi_{\mathsf{SL}}(d)) + [\![G]\!](d) + \Phi(d) \cdot \theta\big)$$

$$= \pi_{\mathsf{TL}}\big(\operatorname*{argmax}_{d} DF(G,s)(d) + \theta_1 \cdot F_1(d) + \cdots + \theta_n \cdot F_n(d)\big)$$

derivation forest for s (weighted tree automaton)

$$DF(G,s) \colon D(G) \to \mathbb{R} \colon d \mapsto [\![(= s)]\!](\pi_{\mathsf{SL}}(d)) + [\![G]\!](d)$$

- CYK+ or Earley algorithm $O(|G| \cdot |s|^3)$

Conventional Model and Decoder POLISHED!

$$
\begin{aligned}
h_{G,\Phi,\theta} \colon s \mapsto \pi_{\mathsf{TL}} \Big(& \operatorname*{argmax}_{d \colon \pi_{\mathsf{SL}}(d) = s} [\![G]\!](d) + \Phi(d) \cdot \theta \Big) \\
= \; & \pi_{\mathsf{TL}} \big(\operatorname*{argmax}_{d} [\![(= s)]\!](\pi_{\mathsf{SL}}(d)) + [\![G]\!](d) + \Phi(d) \cdot \theta \big) \\
= \; & \pi_{\mathsf{TL}} \big(\operatorname*{argmax}_{d} DF(G, s)(d) + \theta_1 \cdot F_1(d) + \cdots + \theta_n \cdot F_n(d) \big) \\
= \; & \pi_{\mathsf{TL}} \big(\operatorname*{argmax}_{d} (DF(G, s) + \theta_1 \cdot F_1 + \cdots + \theta_n \cdot F_n)(d) \big)
\end{aligned}
$$

derivation forest for s (weighted tree automaton)

$$DF(G, s) \colon D(G) \to \mathbb{R} \colon d \mapsto [\![(= s)]\!](\pi_{\mathsf{SL}}(d)) + [\![G]\!](d)$$

- CYK+ or Earley algorithm $O(|G| \cdot |s|^3)$
- Hadamard product of weighted tree automata $O(|A_1| \cdot |A_2|)$

Conventional Model and Decoder POLISHED!

$$
\begin{aligned}
h_{G,\Phi,\theta} \colon s \mapsto \pi_{\mathsf{TL}} \Big(& \operatorname*{argmax}_{d \colon \pi_{\mathsf{SL}}(d) = s} [\![G]\!](d) + \Phi(d) \cdot \theta \Big) \\
= \; & \pi_{\mathsf{TL}} \big(\operatorname*{argmax}_{d} [\![(= s)]\!](\pi_{\mathsf{SL}}(d)) + [\![G]\!](d) + \Phi(d) \cdot \theta \big) \\
= \; & \pi_{\mathsf{TL}} \big(\operatorname*{argmax}_{d} DF(G, s)(d) + \theta_1 \cdot F_1(d) + \cdots + \theta_n \cdot F_n(d) \big) \\
= \; & \pi_{\mathsf{TL}} \big(\operatorname*{argmax}_{d} (DF(G, s) + \theta_1 \cdot F_1 + \cdots + \theta_n \cdot F_n)(d) \big)
\end{aligned}
$$

derivation forest for s (weighted tree automaton)

$$DF(G, s) \colon D(G) \to \mathbb{R} \colon d \mapsto [\![(= s)]\!](\pi_{\mathsf{SL}}(d)) + [\![G]\!](d)$$

- CYK+ or Earley algorithm $O(|G| \cdot |s|^3)$
- Hadamard product of weighted tree automata $O(|A_1| \cdot |A_2|)$
- Knuth algorithm $O(|E| \cdot \log |V|)$

Conventional Model and Decoder POLISHED!

$$h_{G,\Phi,\theta} \colon s \mapsto \pi_{\mathsf{TL}}\Big(\operatorname*{argmax}_{d \colon \pi_{\mathsf{SL}}(d)=s} [\![G]\!](d) + \Phi(d) \cdot \theta\Big)$$

$$= \pi_{\mathsf{TL}}\big(\operatorname*{argmax}_{d} [\![(= s)]\!](\pi_{\mathsf{SL}}(d)) + [\![G]\!](d) + \Phi(d) \cdot \theta\big)$$

$$= \pi_{\mathsf{TL}}\big(\operatorname*{argmax}_{d} DF(G,s)(d) + \theta_1 \cdot F_1(d) + \cdots + \theta_n \cdot F_n(d)\big)$$

$$= \pi_{\mathsf{TL}}\big(\operatorname*{argmax}_{d} (DF(G,s) + \theta_1 \cdot F_1 + \cdots + \theta_n \cdot F_n)(d)\big)$$

derivation forest for s (weighted tree automaton)

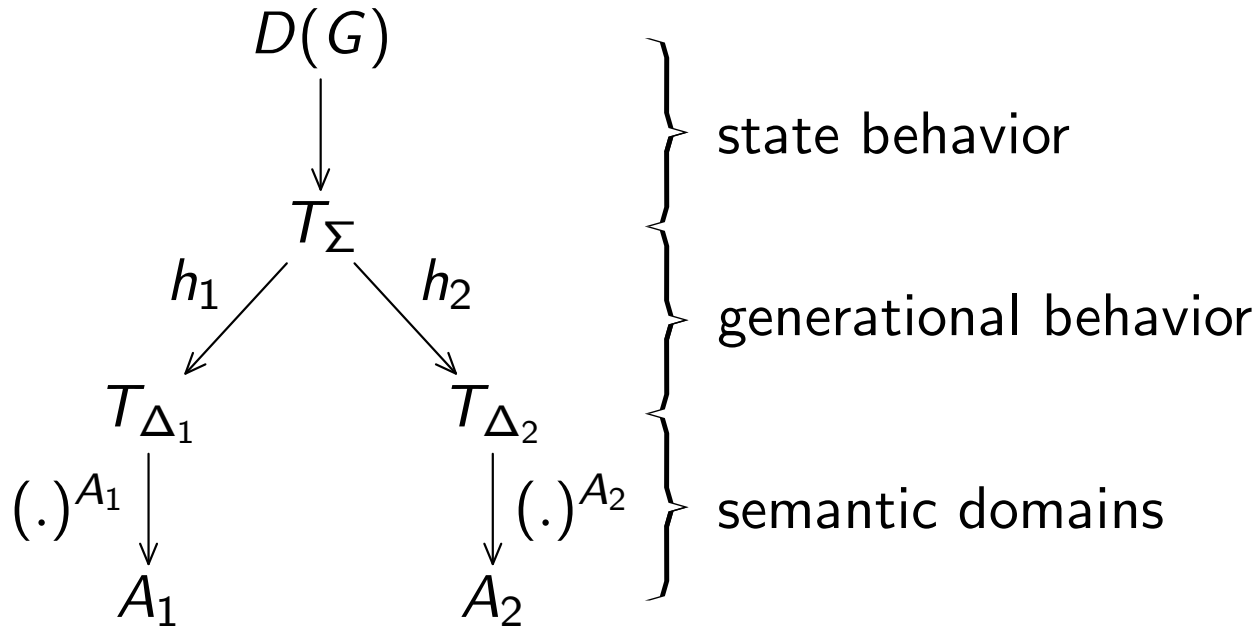
$$DF(G,s) \colon D(G) \to \mathbb{R} \colon d \mapsto [\![(= s)]\!](\pi_{\mathsf{SL}}(d)) + [\![G]\!](d)$$

- CYK+ or Earley algorithm $O(|G| \cdot |s|^3)$
- Hadamard product of weighted tree automata $O(|A_1| \cdot |A_2|)$
- Knuth algorithm $O(|E| \cdot \log|V|)$

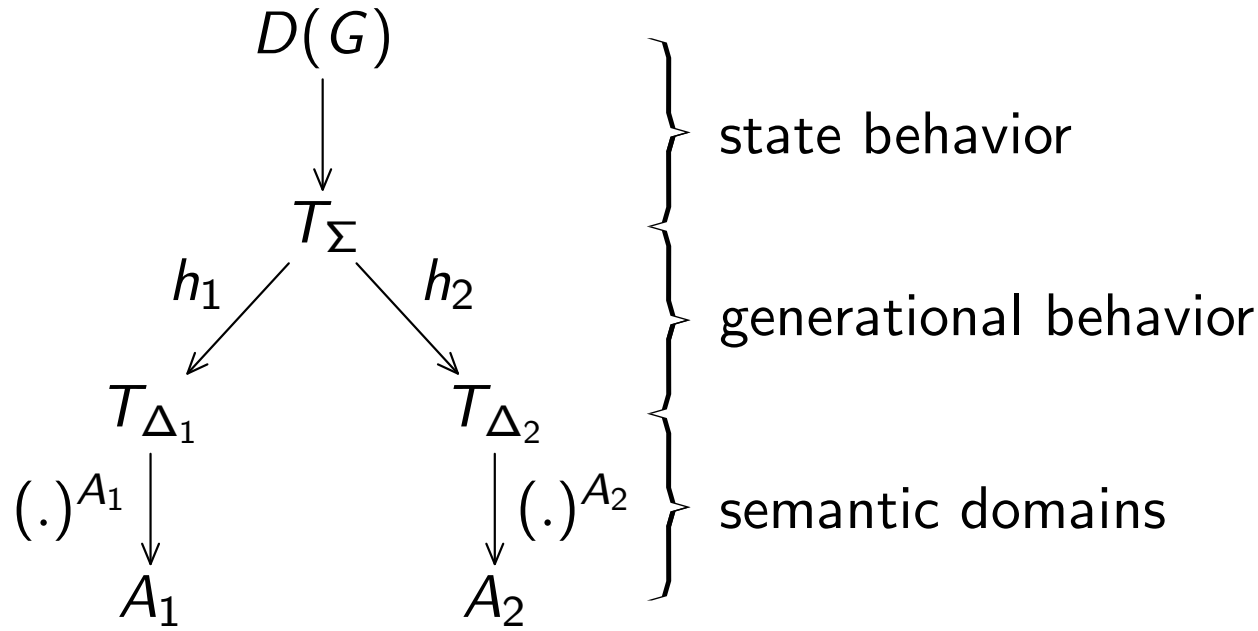
neither program code nor concepts can be reused

Outline

IRTGs over (A_1, A_2) (Koller and Kuhlmann 2011)



IRTGs over (A_1, A_2) (Koller and Kuhlmann 2011)



SCFG rule	$\mathrm{X} \to \langle \text{yu } \mathrm{X}_{\boxed{1}} \text{ you } \mathrm{X}_{\boxed{2}}, \text{ have } \mathrm{X}_{\boxed{2}} \text{ with } \mathrm{X}_{\boxed{1}} \rangle$
RTG rule	$\mathrm{X} \to \sigma(X, X)$
tree hom. h_1	$\sigma \mapsto \mathrm{concat}_4(\text{yu}, x_1, \text{you}, x_2)$
tree hom. h_2	$\sigma \mapsto \mathrm{concat}_4(\text{have}, x_2, \text{with}, x_1)$
A_1, A_2	string algebra: constants, concatenation

Model and Decoder for IRTGs over (A_1, A_2)

$h_{G,\Phi,\theta}$:

$$s \mapsto \pi_{\mathsf{TL}}\left(\mathrm{argmax}_{d:\ \pi_{\mathsf{SL}}(d)=s}\ [\![G]\!](d) + \Phi(d)\cdot\theta\right)$$

Model and Decoder for IRTGs over (A_1, A_2)

$h_{G,\Phi,\theta}$:

$$s \mapsto \pi_{\mathrm{TL}}\left(\mathrm{argmax}_{d:\ \pi_{\mathrm{SL}}(d)=s}\ [\![G]\!](d) + \Phi(d)\cdot\theta\right)$$
$$= \pi_{\mathrm{TL}}\left(\mathrm{argmax}_{d}\ \mathrm{terms}(s)(h_1(d)) + [\![G]\!](d) + \Phi(d)\cdot\theta\right)$$

$$\mathrm{terms}(s)(t) = \begin{cases} 0 & \text{if } t^{A_1} = s \\ -\infty & \text{otherwise} \end{cases}$$

recognizable weighted tree language

Model and Decoder for IRTGs over (A_1, A_2)

$h_{G,\Phi,\theta}$:

$$s \mapsto \pi_{\mathsf{TL}}\left(\mathrm{argmax}_{d\colon \pi_{\mathsf{SL}}(d)=s} [\![G]\!](d) + \Phi(d)\cdot\theta\right)$$
$$= \pi_{\mathsf{TL}}\left(\mathrm{argmax}_{d}\ \mathrm{terms}(s)(h_1(d)) + [\![G]\!](d) + \Phi(d)\cdot\theta\right)$$
$$= \pi_{\mathsf{TL}}\left(\mathrm{argmax}_{d}\ [\![\mathrm{terms}(s) \triangleleft G]\!](d) + \Phi(d)\cdot\theta\right)$$

$$\mathrm{terms}(s)(t) = \begin{cases} 0 & \text{if } t^{A_1} = s \\ -\infty & \text{otherwise} \end{cases}$$

recognizable weighted tree language

input product $[\![L \triangleleft G]\!](d) = L(h_1(d)) + [\![G]\!](d)$

Model and Decoder for IRTGs over (A_1, A_2)

$h_{G,\Phi,\theta}$:

$$
\begin{aligned}
s \mapsto\ & \pi_{\mathsf{TL}}\left(\mathsf{argmax}_{d\colon \pi_{\mathsf{SL}}(d)=s}\ [\![G]\!](d) + \Phi(d)\cdot\theta\right) \\
=\ & \pi_{\mathsf{TL}}\left(\mathsf{argmax}_{d}\ \ \mathsf{terms}(s)(h_1(d)) + [\![G]\!](d) + \Phi(d)\cdot\theta\right) \\
=\ & \pi_{\mathsf{TL}}\left(\mathsf{argmax}_{d}\ [\![\mathsf{terms}(s) \lhd G]\!](d) + \Phi(d)\cdot\theta\right) \\
=\ & \pi_{\mathsf{TL}}\left(\mathsf{argmax}_{d}\ [\![\mathsf{terms}(s) \lhd G]\!](d) + \theta_1\cdot F_1(h_2(d)) + \ldots + \theta_n\cdot F_n(h_2(d))\right)
\end{aligned}
$$

$$
\mathsf{terms}(s)(t) = \begin{cases} 0 & \text{if } t^{A_1} = s \\ -\infty & \text{otherwise} \end{cases}
$$

recognizable weighted tree language

input product $[\![L \lhd G]\!](d) = L(h_1(d)) + [\![G]\!](d)$

Model and Decoder for IRTGs over (A_1, A_2)

$h_{G,\Phi,\theta}$:

$$
\begin{aligned}
s \mapsto\ & \pi_{\mathrm{TL}}\left(\mathrm{argmax}_{d\colon \pi_{\mathrm{SL}}(d)=s}\ [\![G]\!](d) + \Phi(d)\cdot\theta\right) \\
=\ & \pi_{\mathrm{TL}}\left(\mathrm{argmax}_{d}\ \ \mathrm{terms}(s)(h_1(d)) + [\![G]\!](d) + \Phi(d)\cdot\theta\right) \\
=\ & \pi_{\mathrm{TL}}\left(\mathrm{argmax}_{d}\ [\![\mathrm{terms}(s) \triangleleft G]\!](d) + \Phi(d)\cdot\theta\right) \\
=\ & \pi_{\mathrm{TL}}\left(\mathrm{argmax}_{d}\ [\![\mathrm{terms}(s) \triangleleft G]\!](d) + \theta_1\cdot F_1(h_2(d)) + \ldots + \theta_n\cdot F_n(h_2(d))\right) \\
=\ & \pi_{\mathrm{TL}}\left(\mathrm{argmax}_{d}\ [\![\mathrm{terms}(s) \triangleleft G]\!](d) + (\theta_1\cdot F_1 + \ldots + \theta_n\cdot F_n)(h_2(d))\right)
\end{aligned}
$$

$$
\mathrm{terms}(s)(t) = \begin{cases} 0 & \text{if } t^{A_1} = s \\ -\infty & \text{otherwise} \end{cases}
$$

recognizable weighted tree language

input product $[\![L \triangleleft G]\!](d) = L(h_1(d)) + [\![G]\!](d)$

Model and Decoder for IRTGs over (A_1, A_2)

$h_{G,\Phi,\theta}$:

$$
\begin{aligned}
s \mapsto\ & \pi_{\mathsf{TL}}\left(\mathrm{argmax}_{d\colon \pi_{\mathsf{SL}}(d)=s}\ [\![G]\!](d) + \Phi(d)\cdot\theta\right) \\
=\ & \pi_{\mathsf{TL}}\left(\mathrm{argmax}_{d}\ \ \mathrm{terms}(s)(h_1(d)) + [\![G]\!](d) + \Phi(d)\cdot\theta\right) \\
=\ & \pi_{\mathsf{TL}}\left(\mathrm{argmax}_{d}\ [\![\mathrm{terms}(s) \lhd G]\!](d) + \Phi(d)\cdot\theta\right) \\
=\ & \pi_{\mathsf{TL}}\left(\mathrm{argmax}_{d}\ [\![\mathrm{terms}(s) \lhd G]\!](d) + \theta_1\cdot F_1(h_2(d)) + \ldots + \theta_n\cdot F_n(h_2(d))\right) \\
=\ & \pi_{\mathsf{TL}}\left(\mathrm{argmax}_{d}\ [\![\mathrm{terms}(s) \lhd G]\!](d) + (\theta_1\cdot F_1 + \ldots + \theta_n\cdot F_n)(h_2(d))\right) \\
=\ & \pi_{\mathsf{TL}}\left(\mathrm{argmax}_{d}\ [\![\mathrm{terms}(s) \lhd G \rhd (\theta_1\cdot F_1 + \ldots + \theta_n\cdot F_n)]\!](d)\right)
\end{aligned}
$$

$$
\mathrm{terms}(s)(t) = \begin{cases} 0 & \text{if } t^{A_1} = s \\ -\infty & \text{otherwise} \end{cases}
$$

recognizable weighted tree language

input product $[\![L \lhd G]\!](d) = L(h_1(d)) + [\![G]\!](d)$

output product $[\![G \rhd L]\!](d) = [\![G]\!](d) + L(h_2(d))$

Implementation

- IRTG over (strings, trees)
- components in Vanda Toolbox (Haskell)
  - input product (Earley's algorithm)
  - binarization
  - Knuth's algorithm
- accessible in Vanda Studio

References

Chiang, David et al. (2005). "The Hiero machine translation system: extensions, evaluation, and analysis". In: Proc. Human Language Technology and Empirical Methods in Natural Language Processing. ACL, pp. 779–786.

Dyer, Chris et al. (2010). "cdec: A Decoder, Alignment, and Learning Framework for Finite-State and Context-Free Translation Models". In: Proc. ACL 2010 System Demonstrations. ACL, pp. 7–12.

Koehn, Philipp et al. (2007). "Moses: open source toolkit for statistical machine translation". In: Proc. ACL on Interactive Poster and Demonstration Sessions. ACL, pp. 177–180.

Koller, Alexander and Marco Kuhlmann (2011). "A Generalized View on Parsing and Translation". In: Proc. IWPT 2011, pp. 2–13.

Li, Zhifei et al. (2009). "Joshua: an open source toolkit for parsing-based machine translation". In: Proc. Workshop on Statistical Machine Translation. ACL, pp. 135–139.