

Vanda

A Statistical Machine Translation Toolkit

Matthias BÜchse Toni Dietze Johannes Osterholzer
Anja Fischer Linda Leuschner



WATA, Dresden, 2012-05-30

Outline

State of the Art

Vanda: Aims and Solutions

Outline

State of the Art

Vanda: Aims and Solutions

Machine Translation

⋮

ich säge ihre ente

ich sah, wie sie sich duckte

ich esse spaghetti mit der gabel

ich esse spaghetti mit fleischklößen

⋮

F

⋮

i saw her duck

i saw her ducking

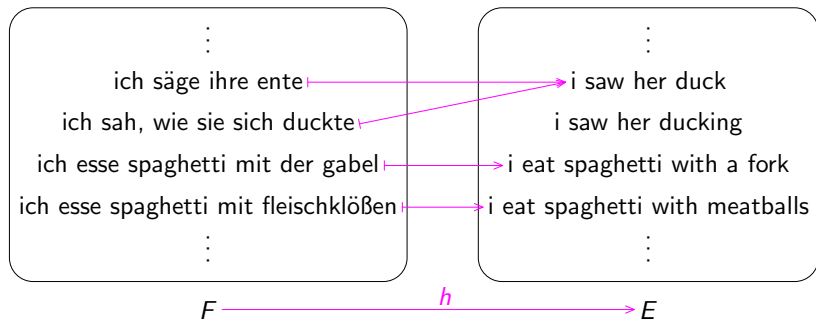
i eat spaghetti with a fork

i eat spaghetti with meatballs

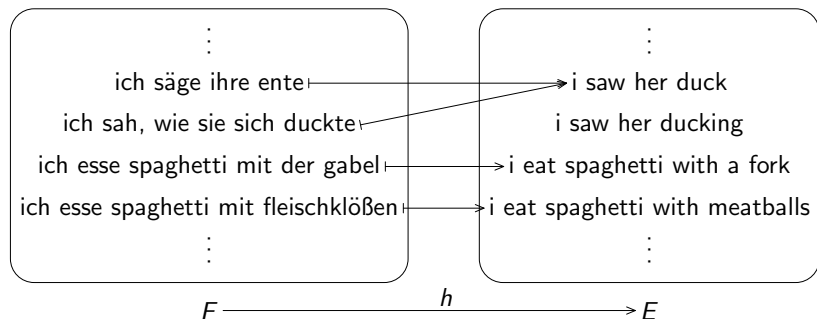
⋮

E

Machine Translation

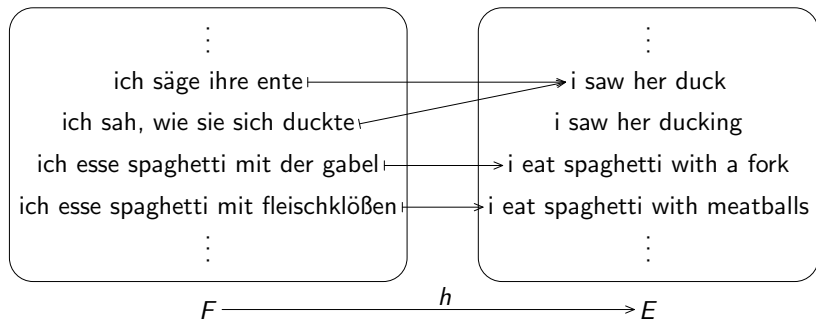


Machine Translation



modelling select $\mathcal{H} \subseteq E^F$

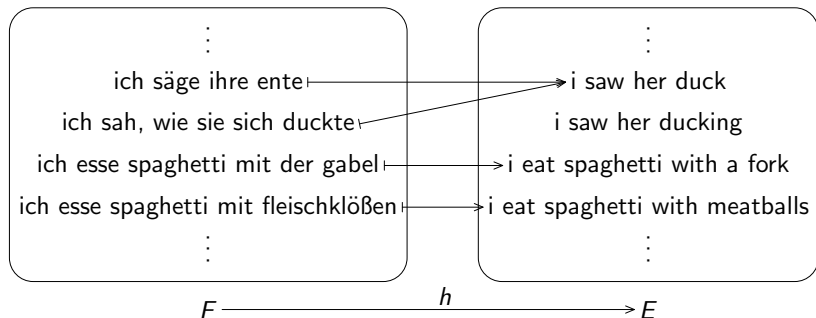
Machine Translation



modelling select $\mathcal{H} \subseteq E^F$

decoding apply any given $h \in \mathcal{H}$

Machine Translation

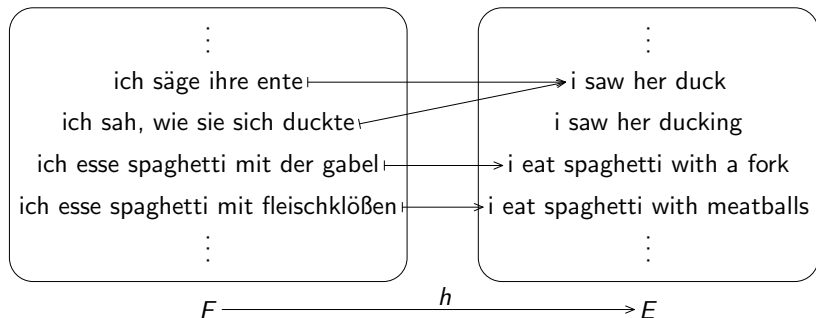


modelling select $\mathcal{H} \subseteq E^F$

decoding apply any given $h \in \mathcal{H}$

training select $h \in \mathcal{H}$

Machine Translation



modelling select $\mathcal{H} \subseteq E^F$

decoding apply any given $h \in \mathcal{H}$

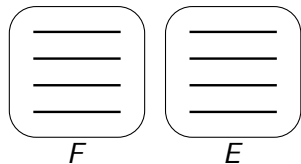
training select $h \in \mathcal{H}$

Modelling: State Of The Art

$$\mathcal{H} = \{h_{G,LM,\theta} \mid G \in \mathcal{G}, LM \in \mathcal{LM}, \theta \in \mathbb{R}^2\}$$

where

$$h_{G,LM,\theta}: f \mapsto$$

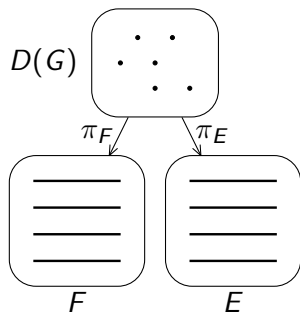


Modelling: State Of The Art

$$\mathcal{H} = \{h_{G,LM,\theta} \mid G \in \mathcal{G}, LM \in \mathcal{LM}, \theta \in \mathbb{R}^2\}$$

where

$$h_{G,LM,\theta}: f \mapsto$$

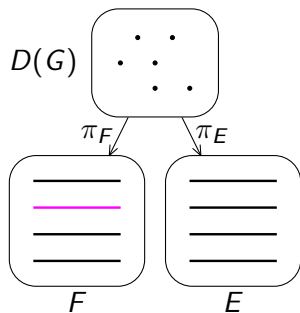


Modelling: State Of The Art

$$\mathcal{H} = \{h_{G,LM,\theta} \mid G \in \mathcal{G}, LM \in \mathcal{LM}, \theta \in \mathbb{R}^2\}$$

where

$$h_{G,LM,\theta}: f \mapsto$$

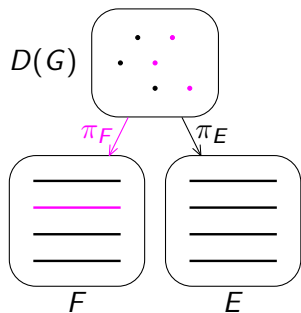


Modelling: State Of The Art

$$\mathcal{H} = \{h_{G,LM,\theta} \mid G \in \mathcal{G}, LM \in \mathcal{LM}, \theta \in \mathbb{R}^2\}$$

where

$$h_{G,LM,\theta}: f \mapsto$$

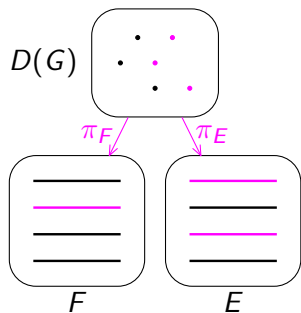


Modelling: State Of The Art

$$\mathcal{H} = \{h_{G,LM,\theta} \mid G \in \mathcal{G}, LM \in \mathcal{LM}, \theta \in \mathbb{R}^2\}$$

where

$$h_{G,LM,\theta}: f \mapsto \pi_E \left(\right)$$

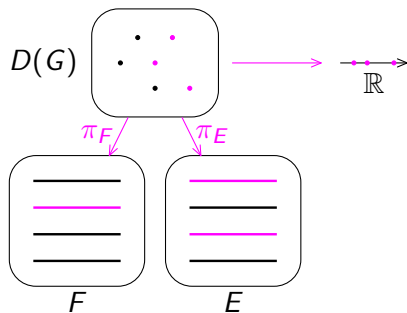


Modelling: State Of The Art

$$\mathcal{H} = \{h_{G,LM,\theta} \mid G \in \mathcal{G}, LM \in \mathcal{LM}, \theta \in \mathbb{R}^2\}$$

where

$$h_{G,LM,\theta}: f \mapsto \pi_E \left(\llbracket G \rrbracket(d)^{\theta_1} \cdot \llbracket LM \rrbracket(\pi_E(d))^{\theta_2} \right)$$

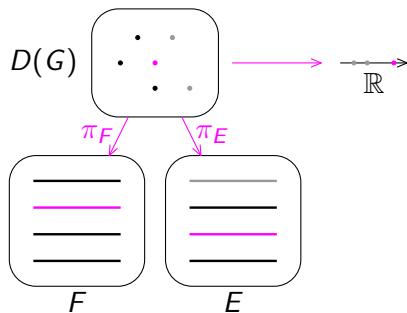


Modelling: State Of The Art

$$\mathcal{H} = \{h_{G,LM,\theta} \mid G \in \mathcal{G}, LM \in \mathcal{LM}, \theta \in \mathbb{R}^2\}$$

where

$$h_{G,LM,\theta}: f \mapsto \pi_E \left(\underset{d: \pi_F(d)=f}{\operatorname{argmax}} \llbracket G \rrbracket (d)^{\theta_1} \cdot \llbracket LM \rrbracket (\pi_E(d))^{\theta_2} \right)$$



Synchronous Context-Free Grammar

$\pi_1: S \rightarrow \langle S X, S X \rangle$

$\pi_2: S \rightarrow \langle X, X \rangle$

$\pi_3: X \rightarrow \langle \text{yu } X_{[1]} \text{ you } X_{[2]}, \text{ have } X_{[2]} \text{ with } X_{[1]} \rangle$

$\pi_4: X \rightarrow \langle X_{[1]} \text{ de } X_{[2]}, \text{ the } X_{[2]} \text{ that } X_{[1]} \rangle$

$\pi_5: X \rightarrow \langle X \text{ zhiyi, one of } X \rangle$

$\pi_6: X \rightarrow \langle \text{Aozhou, Australia} \rangle$

$\pi_7: X \rightarrow \langle \text{Beihan, North Korea} \rangle$

$\pi_8: X \rightarrow \langle \text{shi, is} \rangle$

$\pi_9: X \rightarrow \langle \text{bangjiao, diplomatic relations} \rangle$

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

Derivation

$\langle S, S \rangle$

Derivation

$$\begin{array}{l} \langle S, S \rangle \\ \xRightarrow{\pi_1} \langle S X, S X \rangle \end{array}$$

Derivation

$$\begin{array}{l} \langle S, S \rangle \\ \xRightarrow{\pi_1} \langle S X, S X \rangle \end{array}$$

Derivation

$$\begin{aligned} & \langle S, S \rangle \\ \xRightarrow{\pi_1} & \langle S X, S X \rangle \\ \xRightarrow{\pi_1} & \langle S X_{\boxed{1}} X_{\boxed{2}}, S X_{\boxed{1}} X_{\boxed{2}} \rangle \end{aligned}$$

Derivation

$$\begin{aligned} & \langle S, S \rangle \\ \xRightarrow{\pi_1} & \langle S X, S X \rangle \\ \xRightarrow{\pi_1} & \langle S X_{\boxed{1}} X_{\boxed{2}}, S X_{\boxed{1}} X_{\boxed{2}} \rangle \end{aligned}$$

Derivation

$$\begin{aligned} & \langle S, S \rangle \\ \xRightarrow{\pi_1} & \langle S X, S X \rangle \\ \xRightarrow{\pi_1} & \langle S X_{\boxed{1}} X_{\boxed{2}}, S X_{\boxed{1}} X_{\boxed{2}} \rangle \\ \xRightarrow{\pi_2} & \langle X_{\boxed{0}} X_{\boxed{1}} X_{\boxed{2}}, X_{\boxed{0}} X_{\boxed{1}} X_{\boxed{2}} \rangle \end{aligned}$$

Derivation

$$\begin{aligned} & \langle S, S \rangle \\ \xRightarrow{\pi_1} & \langle S X, S X \rangle \\ \xRightarrow{\pi_1} & \langle S X_{\boxed{1}} X_{\boxed{2}}, S X_{\boxed{1}} X_{\boxed{2}} \rangle \\ \xRightarrow{\pi_2} & \langle X_{\boxed{0}} X_{\boxed{1}} X_{\boxed{2}}, X_{\boxed{0}} X_{\boxed{1}} X_{\boxed{2}} \rangle \end{aligned}$$

Derivation

$$\begin{aligned} & \langle S, S \rangle \\ \xRightarrow{\pi_1} & \langle S X, S X \rangle \\ \xRightarrow{\pi_1} & \langle S X_{\boxed{1}} X_{\boxed{2}}, S X_{\boxed{1}} X_{\boxed{2}} \rangle \\ \xRightarrow{\pi_2} & \langle X_{\boxed{0}} X_{\boxed{1}} X_{\boxed{2}}, X_{\boxed{0}} X_{\boxed{1}} X_{\boxed{2}} \rangle \\ \xRightarrow{\pi_6} & \langle \text{Aozhou } X_{\boxed{1}} X_{\boxed{2}}, \text{Australia } X_{\boxed{1}} X_{\boxed{2}} \rangle \end{aligned}$$

Derivation

$\langle S, S \rangle$

$\xRightarrow{\pi_1} \langle S X, S X \rangle$

$\xRightarrow{\pi_1} \langle S X_{[1]} X_{[2]}, S X_{[1]} X_{[2]} \rangle$

$\xRightarrow{\pi_2} \langle X_{[0]} X_{[1]} X_{[2]}, X_{[0]} X_{[1]} X_{[2]} \rangle$

$\xRightarrow{\pi_6} \langle \text{Aozhou } X_{[1]} X_{[2]}, \text{Australia } X_{[1]} X_{[2]} \rangle$

Derivation

$\langle S, S \rangle$

$\xRightarrow{\pi_1} \langle S X, S X \rangle$

$\xRightarrow{\pi_1} \langle S X_{[1]} X_{[2]}, S X_{[1]} X_{[2]} \rangle$

$\xRightarrow{\pi_2} \langle X_{[0]} X_{[1]} X_{[2]}, X_{[0]} X_{[1]} X_{[2]} \rangle$

$\xRightarrow{\pi_6} \langle \text{Aozhou } X_{[1]} X_{[2]}, \text{Australia } X_{[1]} X_{[2]} \rangle$

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

Derivation

$\langle S, S \rangle$

$\xRightarrow{\pi_1} \langle S X, S X \rangle$

$\xRightarrow{\pi_1} \langle S X_{[1]} X_{[2]}, S X_{[1]} X_{[2]} \rangle$

$\xRightarrow{\pi_2} \langle X_{[0]} X_{[1]} X_{[2]}, X_{[0]} X_{[1]} X_{[2]} \rangle$

$\xRightarrow{\pi_6} \langle \text{Aozhou } X_{[1]} X_{[2]}, \text{Australia } X_{[1]} X_{[2]} \rangle$

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

Derivation

$\langle S, S \rangle$
 $\xRightarrow{\pi_1} \langle S X, S X \rangle$
 $\xRightarrow{\pi_1} \langle S X_{[1]} X_{[2]}, S X_{[1]} X_{[2]} \rangle$
 $\xRightarrow{\pi_2} \langle X_{[0]} X_{[1]} X_{[2]}, X_{[0]} X_{[1]} X_{[2]} \rangle$
 $\xRightarrow{\pi_6} \langle \text{Aozhou } X_{[1]} X_{[2]}, \text{Australia } X_{[1]} X_{[2]} \rangle$
 $\xRightarrow{\pi_8} \langle \text{Aozhou shi } X, \text{Australia is } X \rangle$
 $\xRightarrow{\pi_5} \langle \text{Aozhou shi } X \text{ zhiyi}, \text{Australia is one of } X \rangle$

Derivation

$\langle S, S \rangle$

$\xRightarrow{\pi_1} \langle S X, S X \rangle$

$\xRightarrow{\pi_1} \langle S X_{[1]} X_{[2]}, S X_{[1]} X_{[2]} \rangle$

$\xRightarrow{\pi_2} \langle X_{[0]} X_{[1]} X_{[2]}, X_{[0]} X_{[1]} X_{[2]} \rangle$

$\xRightarrow{\pi_6} \langle \text{Aozhou } X_{[1]} X_{[2]}, \text{Australia } X_{[1]} X_{[2]} \rangle$

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

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

Derivation

$\langle S, S \rangle$

$\xRightarrow{\pi_1} \langle S X, S X \rangle$

$\xRightarrow{\pi_1} \langle S X_{[1]} X_{[2]}, S X_{[1]} X_{[2]} \rangle$

$\xRightarrow{\pi_2} \langle X_{[0]} X_{[1]} X_{[2]}, X_{[0]} X_{[1]} X_{[2]} \rangle$

$\xRightarrow{\pi_6} \langle \text{Aozhou } X_{[1]} X_{[2]}, \text{Australia } X_{[1]} X_{[2]} \rangle$

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

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

$\xRightarrow{\pi_4} \langle \text{Aozhou shi } X_{[1]} \text{ de } X_{[2]} \text{ zhiyi,}$
 $\text{Australia is one of the } X_{[2]} \text{ that } X_{[1]} \rangle$

Derivation

$\langle S, S \rangle$

$\xRightarrow{\pi_1} \langle S X, S X \rangle$

$\xRightarrow{\pi_1} \langle S X_{[1]} X_{[2]}, S X_{[1]} X_{[2]} \rangle$

$\xRightarrow{\pi_2} \langle X_{[0]} X_{[1]} X_{[2]}, X_{[0]} X_{[1]} X_{[2]} \rangle$

$\xRightarrow{\pi_6} \langle \text{Aozhou } X_{[1]} X_{[2]}, \text{Australia } X_{[1]} X_{[2]} \rangle$

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

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

$\xRightarrow{\pi_4} \langle \text{Aozhou shi } X_{[1]} \text{ de } X_{[2]} \text{ zhiyi,}$
 $\text{Australia is one of the } X_{[2]} \text{ that } X_{[1]} \rangle$

Derivation

$\langle S, S \rangle$

$\xRightarrow{\pi_1} \langle S X, S X \rangle$

$\xRightarrow{\pi_1} \langle S X_{[1]} X_{[2]}, S X_{[1]} X_{[2]} \rangle$

$\xRightarrow{\pi_2} \langle X_{[0]} X_{[1]} X_{[2]}, X_{[0]} X_{[1]} X_{[2]} \rangle$

$\xRightarrow{\pi_6} \langle \text{Aozhou } X_{[1]} X_{[2]}, \text{Australia } X_{[1]} X_{[2]} \rangle$

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

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

$\xRightarrow{\pi_4} \langle \text{Aozhou shi } X_{[1]} \text{ de } X_{[2]} \text{ zhiyi,}$
 $\text{Australia is one of the } X_{[2]} \text{ that } X_{[1]} \rangle$

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

Derivation

$\langle S, S \rangle$

$\xRightarrow{\pi_1} \langle S X, S X \rangle$

$\xRightarrow{\pi_1} \langle S X_{[1]} X_{[2]}, S X_{[1]} X_{[2]} \rangle$

$\xRightarrow{\pi_2} \langle X_{[0]} X_{[1]} X_{[2]}, X_{[0]} X_{[1]} X_{[2]} \rangle$

$\xRightarrow{\pi_6} \langle \text{Aozhou } X_{[1]} X_{[2]}, \text{Australia } X_{[1]} X_{[2]} \rangle$

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

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

$\xRightarrow{\pi_4} \langle \text{Aozhou shi } X_{[1]} \text{ de } X_{[2]} \text{ zhiyi},$
 $\text{Australia is one of the } X_{[2]} \text{ that } X_{[1]} \rangle$

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

Derivation

$\langle S, S \rangle$

$\xRightarrow{\pi_1} \langle S X, S X \rangle$

$\xRightarrow{\pi_1} \langle S X_{[1]} X_{[2]}, S X_{[1]} X_{[2]} \rangle$

$\xRightarrow{\pi_2} \langle X_{[0]} X_{[1]} X_{[2]}, X_{[0]} X_{[1]} X_{[2]} \rangle$

$\xRightarrow{\pi_6} \langle \text{Aozhou } X_{[1]} X_{[2]}, \text{Australia } X_{[1]} X_{[2]} \rangle$

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

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

$\xRightarrow{\pi_4} \langle \text{Aozhou shi } X_{[1]} \text{ de } X_{[2]} \text{ zhiyi},$
 $\text{Australia is one of the } X_{[2]} \text{ that } X_{[1]} \rangle$

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

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

Derivation

$\langle S, S \rangle$

$\xRightarrow{\pi_1} \langle S X, S X \rangle$

$\xRightarrow{\pi_1} \langle S X_{[1]} X_{[2]}, S X_{[1]} X_{[2]} \rangle$

$\xRightarrow{\pi_2} \langle X_{[0]} X_{[1]} X_{[2]}, X_{[0]} X_{[1]} X_{[2]} \rangle$

$\xRightarrow{\pi_6} \langle \text{Aozhou } X_{[1]} X_{[2]}, \text{Australia } X_{[1]} X_{[2]} \rangle$

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

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

$\xRightarrow{\pi_4} \langle \text{Aozhou shi } X_{[1]} \text{ de } X_{[2]} \text{ zhiyi,}$
 $\text{Australia is one of the } X_{[2]} \text{ that } X_{[1]} \rangle$

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

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

Derivation

$\langle S, S \rangle$

$\xRightarrow{\pi_1} \langle S X, S X \rangle$

$\xRightarrow{\pi_1} \langle S X_{[1]} X_{[2]}, S X_{[1]} X_{[2]} \rangle$

$\xRightarrow{\pi_2} \langle X_{[0]} X_{[1]} X_{[2]}, X_{[0]} X_{[1]} X_{[2]} \rangle$

$\xRightarrow{\pi_6} \langle \text{Aozhou } X_{[1]} X_{[2]}, \text{Australia } X_{[1]} X_{[2]} \rangle$

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

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

$\xRightarrow{\pi_4} \langle \text{Aozhou shi } X_{[1]} \text{ de } X_{[2]} \text{ zhiyi},$
 $\text{Australia is one of the } X_{[2]} \text{ that } X_{[1]} \rangle$

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

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

$\xRightarrow{\pi_9} \langle \text{Aozhou shi yu Beihan you } \text{bangjiao} \text{ de } X_{[2]} \text{ zhiyi},$
 $\text{Australia is one of the } X_{[2]} \text{ that have } \text{diplomatic relations} \text{ with } \dots \rangle$

Derivation

$\langle S, S \rangle$

$\xRightarrow{\pi_1} \langle S X, S X \rangle$

$\xRightarrow{\pi_1} \langle S X_{[1]} X_{[2]}, S X_{[1]} X_{[2]} \rangle$

$\xRightarrow{\pi_2} \langle X_{[0]} X_{[1]} X_{[2]}, X_{[0]} X_{[1]} X_{[2]} \rangle$

$\xRightarrow{\pi_6} \langle \text{Aozhou } X_{[1]} X_{[2]}, \text{Australia } X_{[1]} X_{[2]} \rangle$

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

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

$\xRightarrow{\pi_4} \langle \text{Aozhou shi } X_{[1]} \text{ de } X_{[2]} \text{ zhiyi},$
 $\text{Australia is one of the } X_{[2]} \text{ that } X_{[1]} \rangle$

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

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

$\xRightarrow{\pi_9} \langle \text{Aozhou shi yu Beihan you bangjiao de } X_{[2]} \text{ zhiyi},$
 $\text{Australia is one of the } X_{[2]} \text{ that have diplomatic relations with } \dots \rangle$

Derivation

$\langle S, S \rangle$

$\xRightarrow{\pi_1} \langle S X, S X \rangle$

$\xRightarrow{\pi_1} \langle S X_{[1]} X_{[2]}, S X_{[1]} X_{[2]} \rangle$

$\xRightarrow{\pi_2} \langle X_{[0]} X_{[1]} X_{[2]}, X_{[0]} X_{[1]} X_{[2]} \rangle$

$\xRightarrow{\pi_6} \langle \text{Aozhou } X_{[1]} X_{[2]}, \text{Australia } X_{[1]} X_{[2]} \rangle$

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

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

$\xRightarrow{\pi_4} \langle \text{Aozhou shi } X_{[1]} \text{ de } X_{[2]} \text{ zhiyi},$
 $\text{Australia is one of the } X_{[2]} \text{ that } X_{[1]} \rangle$

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

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

$\xRightarrow{\pi_9} \langle \text{Aozhou shi yu Beihan you bangjiao de } X_{[2]} \text{ zhiyi},$
 $\text{Australia is one of the } X_{[2]} \text{ that have diplomatic relations with } \dots \rangle$

$\xRightarrow{\pi_{10}} \langle \text{Aozhou shi yu Beihan you bangjiao de } \text{shaoshu guojia} \text{ zhiyi},$
 $\text{Australia is one of the } \text{few countries} \text{ that have } \dots \rangle$

Derivation

$\langle S, S \rangle$

$\xRightarrow{\pi_1} \langle S X, S X \rangle$

$\xRightarrow{\pi_1} \langle S X_{[1]} X_{[2]}, S X_{[1]} X_{[2]} \rangle$

$\xRightarrow{\pi_2} \langle X_{[0]} X_{[1]} X_{[2]}, X_{[0]} X_{[1]} X_{[2]} \rangle$

$\xRightarrow{\pi_6} \langle \text{Aozhou } X_{[1]} X_{[2]}, \text{Australia } X_{[1]} X_{[2]} \rangle$

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

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

$\xRightarrow{\pi_4} \langle \text{Aozhou shi } X_{[1]} \text{ de } X_{[2]} \text{ zhiyi},$
 $\text{Australia is one of the } X_{[2]} \text{ that } X_{[1]} \rangle$

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

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

$\xRightarrow{\pi_9} \langle \text{Aozhou shi yu Beihan you bangjiao de } X_{[2]} \text{ zhiyi},$
 $\text{Australia is one of the } X_{[2]} \text{ that have diplomatic relations with } \dots \rangle$

$\xRightarrow{\pi_{10}} \langle \text{Aozhou shi yu Beihan you bangjiao de shaoshu guojia zhiyi},$
 $\text{Australia is one of the few countries that have } \dots \rangle$

Advancing the State of the Art

grammar	implementation	reference
SCFG	Hiero (closed source)	[CLM ⁺ 05]
	Moses (C++)	[KHB ⁺ 07]
	Joshua (Java)	[LCBD ⁺ 09]
	cdec (C++)	[DLG ⁺ 10]

Advancing the State of the Art

grammar	implementation	reference
SCFG	Hiero (closed source)	[CLM ⁺ 05]
	Moses (C++)	[KHB ⁺ 07]
	Joshua (Java)	[LCBD ⁺ 09]
	cdec (C++)	[DLG ⁺ 10]
XTOPs	closed source	[GKM08]
STAG	closed source	[DK09]
MBOT	extension to Moses	[Mal10]

Advancing the State of the Art

grammar	implementation	reference
SCFG	Hiero (closed source)	[CLM ⁺ 05]
	Moses (C++)	[KHB ⁺ 07]
	Joshua (Java)	[LCBD ⁺ 09]
	cdec (C++)	[DLG ⁺ 10]
XTOPs	closed source	[GKM08]
STAG	closed source	[DK09]
MBOT	extension to Moses	[Mal10]

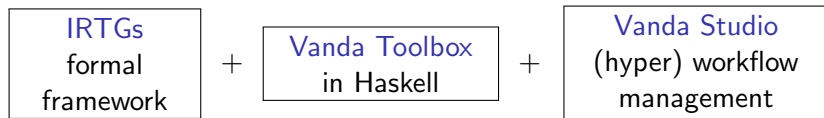
- ▶ monolithic algorithms, specifically implemented
- ▶ little standardisation, manual scripting
- ▶ lack of documentation

Outline

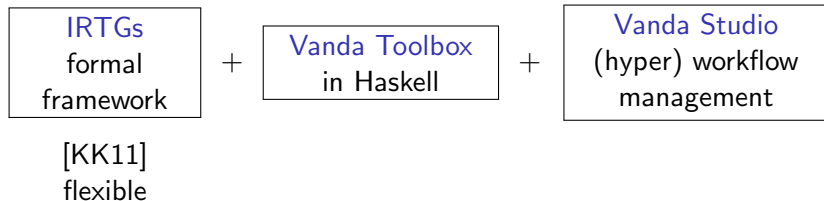
State of the Art

Vanda: Aims and Solutions

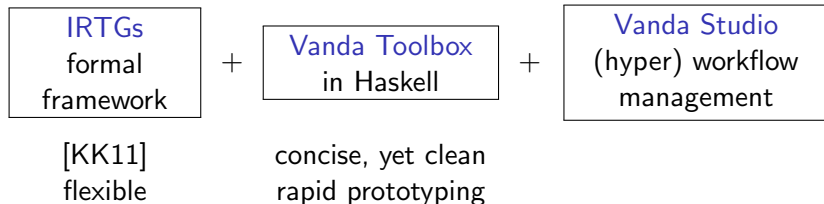
Vanda Components



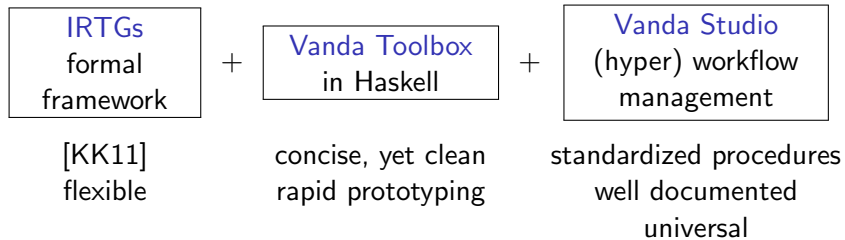
Vanda Components



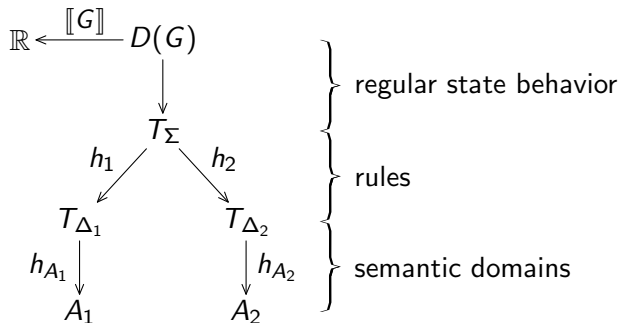
Vanda Components



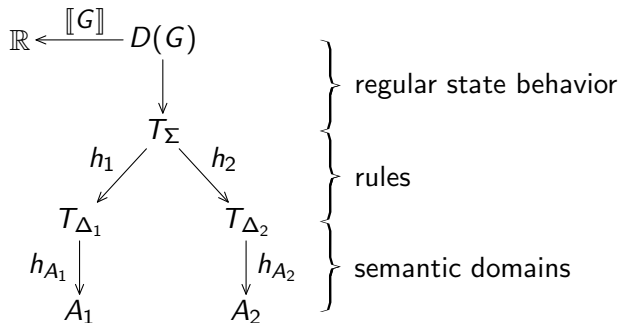
Vanda Components



Formal Framework: IRTGs over (A_1, A_2)



Formal Framework: IRTGs over (A_1, A_2)

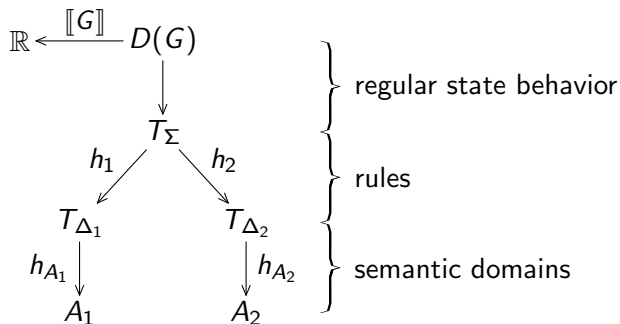


Subsume:

A_1

A_2

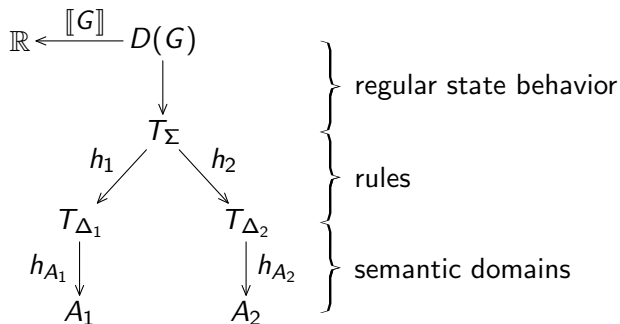
Formal Framework: IRTGs over (A_1, A_2)



Subsume:

	A_1	A_2
SCFG	F , concatenation	E , concatenation

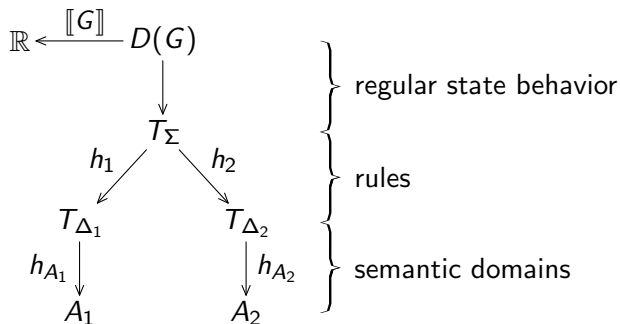
Formal Framework: IRTGs over (A_1, A_2)



Subsume:

	A_1	A_2
InXTOPs	F , concatenation	T_E , top concatenation

Formal Framework: IRTGs over (A_1, A_2)



Subsume:

	A_1	A_2
STAG	T_F , top concatenation, second-order subst.	T_E , top concatenation, second-order subst.

Modularity

$$h_{G,LM,\theta}: f \mapsto \pi_E \left(\operatorname{argmax}_{d: \pi_F(d)=f} \llbracket G \rrbracket(d)^{\theta_1} \cdot \llbracket LM \rrbracket(\pi_E(d))^{\theta_2} \right)$$

Modularity

$$\begin{aligned} h_{G,LM,\theta}: f &\mapsto \pi_E \left(\operatorname{argmax}_{d: \pi_F(d)=f} \llbracket G \rrbracket(d)^{\theta_1} \cdot \llbracket LM \rrbracket(\pi_E(d))^{\theta_2} \right) \\ &= \pi_E \left(\operatorname{argmax}_d \llbracket (=f) \rrbracket(\pi_F(d)) \cdot \llbracket G \rrbracket(d)^{\theta_1} \cdot \llbracket LM \rrbracket(\pi_E(d))^{\theta_2} \right) \end{aligned}$$

Modularity

$$\begin{aligned}h_{G,LM,\theta}: f &\mapsto \pi_E \left(\operatorname{argmax}_{d: \pi_F(d)=f} \llbracket G \rrbracket(d)^{\theta_1} \cdot \llbracket LM \rrbracket(\pi_E(d))^{\theta_2} \right) \\ &= \pi_E \left(\operatorname{argmax}_d \llbracket (= f) \rrbracket(\pi_F(d)) \cdot \llbracket G \rrbracket(d)^{\theta_1} \cdot \llbracket LM \rrbracket(\pi_E(d))^{\theta_2} \right) \\ &= \pi_E \left(\operatorname{argmax}_d (\llbracket (= f) \rrbracket \triangleleft \llbracket G \rrbracket^{\theta_1} \triangleright \llbracket LM \rrbracket^{\theta_2})(d) \right)\end{aligned}$$

Modularity

$$\begin{aligned}h_{G,LM,\theta}: f &\mapsto \pi_E \left(\operatorname{argmax}_{d: \pi_F(d)=f} \llbracket G \rrbracket(d)^{\theta_1} \cdot \llbracket LM \rrbracket(\pi_E(d))^{\theta_2} \right) \\&= \pi_E \left(\operatorname{argmax}_d \llbracket (= f) \rrbracket(\pi_F(d)) \cdot \llbracket G \rrbracket(d)^{\theta_1} \cdot \llbracket LM \rrbracket(\pi_E(d))^{\theta_2} \right) \\&= \pi_E \left(\operatorname{argmax}_d (\llbracket (= f) \rrbracket \triangleleft \llbracket G \rrbracket^{\theta_1} \triangleright \llbracket LM \rrbracket^{\theta_2})(d) \right) \\&= \pi_E \left(\operatorname{argmax}_d \llbracket (= f) \triangleleft G^{\theta_1} \triangleright LM^{\theta_2} \rrbracket(d) \right)\end{aligned}$$

Modularity

$$\begin{aligned}h_{G,LM,\theta}: f &\mapsto \pi_E \left(\operatorname{argmax}_{d: \pi_F(d)=f} \llbracket G \rrbracket(d)^{\theta_1} \cdot \llbracket LM \rrbracket(\pi_E(d))^{\theta_2} \right) \\&= \pi_E \left(\operatorname{argmax}_d \llbracket (=f) \rrbracket(\pi_F(d)) \cdot \llbracket G \rrbracket(d)^{\theta_1} \cdot \llbracket LM \rrbracket(\pi_E(d))^{\theta_2} \right) \\&= \pi_E \left(\operatorname{argmax}_d (\llbracket (=f) \rrbracket \triangleleft \llbracket G \rrbracket^{\theta_1} \triangleright \llbracket LM \rrbracket^{\theta_2})(d) \right) \\&= \pi_E \left(\operatorname{argmax}_d \llbracket (=f) \triangleleft G^{\theta_1} \triangleright LM^{\theta_2} \rrbracket(d) \right)\end{aligned}$$

$$\begin{aligned}\varphi: \operatorname{supp}(\llbracket G \triangleright LM \rrbracket) &\rightarrow \operatorname{supp}(\llbracket G \rrbracket \triangleright \llbracket LM \rrbracket) \\ \llbracket G \triangleright LM \rrbracket &= (\llbracket G \rrbracket \triangleright \llbracket LM \rrbracket) \circ \varphi\end{aligned}$$

Modularity

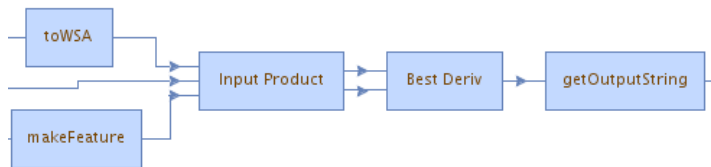
$$\begin{aligned}h_{G,LM,\theta}: f &\mapsto \pi_E \left(\operatorname{argmax}_{d: \pi_F(d)=f} \llbracket G \rrbracket(d)^{\theta_1} \cdot \llbracket LM \rrbracket(\pi_E(d))^{\theta_2} \right) \\&= \pi_E \left(\operatorname{argmax}_d \llbracket (= f) \rrbracket(\pi_F(d)) \cdot \llbracket G \rrbracket(d)^{\theta_1} \cdot \llbracket LM \rrbracket(\pi_E(d))^{\theta_2} \right) \\&= \pi_E \left(\operatorname{argmax}_d (\llbracket (= f) \rrbracket \triangleleft \llbracket G \rrbracket^{\theta_1} \triangleright \llbracket LM \rrbracket^{\theta_2})(d) \right) \\&= \pi_E \left(\operatorname{argmax}_d \llbracket (= f) \triangleleft G^{\theta_1} \triangleright LM^{\theta_2} \rrbracket(d) \right)\end{aligned}$$

$$\begin{aligned}\varphi: \operatorname{supp}(\llbracket G \triangleright LM \rrbracket) &\rightarrow \operatorname{supp}(\llbracket G \rrbracket \triangleright \llbracket LM \rrbracket) \\ \llbracket G \triangleright LM \rrbracket &= (\llbracket G \rrbracket \triangleright \llbracket LM \rrbracket) \circ \varphi\end{aligned}$$

$$\text{RTG} \triangleleft \text{STAG} \subseteq \text{STAG} \supseteq \text{STAG} \triangleright \text{RTG} \quad [\text{BNV11}]$$

Modularity

$$\begin{aligned}h_{G,LM,\theta}: f &\mapsto \pi_E \left(\operatorname{argmax}_{d: \pi_F(d)=f} \llbracket G \rrbracket(d)^{\theta_1} \cdot \llbracket LM \rrbracket(\pi_E(d))^{\theta_2} \right) \\ &= \pi_E \left(\operatorname{argmax}_d \llbracket (= f) \rrbracket(\pi_F(d)) \cdot \llbracket G \rrbracket(d)^{\theta_1} \cdot \llbracket LM \rrbracket(\pi_E(d))^{\theta_2} \right) \\ &= \pi_E \left(\operatorname{argmax}_d (\llbracket (= f) \rrbracket \triangleleft \llbracket G \rrbracket^{\theta_1} \triangleright \llbracket LM \rrbracket^{\theta_2})(d) \right) \\ &= \pi_E \left(\operatorname{argmax}_d \llbracket (= f) \triangleleft G^{\theta_1} \triangleright LM^{\theta_2} \rrbracket(d) \right)\end{aligned}$$



Current State

- ▶ **Vanda Toolbox**: small-scale data; state-split grammars, input product, output product, n-best derivations, KA^* , string-to-tree rule extraction (GHKM), inside/outside EM
- ▶ **Vanda Studio**: proof-of-concept implementation (10K lines), code generation for single workstation

Current State

- ▶ **Vanda Toolbox**: small-scale data; state-split grammars, input product, output product, n-best derivations, KA*, string-to-tree rule extraction (GHKM), inside/outside EM
- ▶ **Vanda Studio**: proof-of-concept implementation (10K lines), code generation for single workstation

I believe that this community should commit itself to achieving the goal, before this decade is out, of providing versatile off-the-shelf components for machine translation and using them successfully in a large-scale task.

References I



Matthias BÜchse, Mark-Jan Nederhof, and Heiko Vogler.
Tree parsing with synchronous tree-adjoining grammars.
In Proceedings of IWPT, 2011.



David Chiang, Adam Lopez, Nitin Madnani, Christof Monz, Philip Resnik, and Michael Subotin.
The Hiero machine translation system: extensions, evaluation, and analysis.
In HLT '05: Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing, pages 779–786, Morristown, NJ, USA, 2005. Association for Computational Linguistics.



Steve DeNeefe and Kevin Knight.
Synchronous tree-adjoining machine translation.
In EMNLP '09: Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing, pages 727–736, Morristown, NJ, USA, 2009. Association for Computational Linguistics.



Chris Dyer, Adam Lopez, Juri Ganitkevitch, Jonathan Weese, Ferhan Ture, Phil Blunsom, Hendra Setiawan, Vladimir Eidelman, and Philip Resnik.
cdec: A decoder, alignment, and learning framework for finite-state and context-free translation models.
In Proceedings of the ACL 2010 System Demonstrations, pages 7–12, Uppsala, Sweden, July 2010. Association for Computational Linguistics.



Jonathan Graehl, Kevin Knight, and Jonathan May.
Training tree transducers.
Comp. Ling., 34(3):391–427, 2008.

References II



Philipp Koehn, Hieu Hoang, Alexandra Birch, Chris Callison-Burch, Marcello Federico, Nicola Bertoldi, Brooke Cowan, Wade Shen, Christine Moran, Richard Zens, Chris Dyer, Ondřej Bojar, Alexandra Constantin, and Evan Herbst.

Moses: open source toolkit for statistical machine translation.

In *Proceedings of the 45th Annual Meeting of the ACL on Interactive Poster and Demonstration Sessions*, ACL '07, pages 177–180, Stroudsburg, PA, USA, 2007. Association for Computational Linguistics.



Alexander Koller and Marco Kuhlmann.

A generalized view on parsing and translation.

In *Proceedings IWPT 2011*, 2011.



Zhifei Li, Chris Callison-Burch, Chris Dyer, Juri Ganitkevitch, Sanjeev Khudanpur, Lane Schwartz, Wren N. G. Thornton, Jonathan Weese, and Omar F. Zaidan.

Joshua: an open source toolkit for parsing-based machine translation.

In *Proceedings of the Fourth Workshop on Statistical Machine Translation*, StatMT '09, pages 135–139, Stroudsburg, PA, USA, 2009. Association for Computational Linguistics.



Andreas Maletti.

Why synchronous tree substitution grammars?

In *Proc. 11th Conf. North American Chapter of the Association for Computational Linguistics*. Association for Computational Linguistics, 2010.