Statistical Machine Translation of Natural Languages

Heiko Vogler
Technische Universität Dresden
Germany

Graduiertenkolleg "Quantitative Logics and Automata"
Dresden, November, 2012

Weighted Tree Automata and Weighted Tree Transducers

can help in

Statistical Machine Translation of Natural Languages

Heiko Vogler
Technische Universität Dresden
Germany

Graduiertenkolleg “Quantitative Logics and Automata”
Dresden, November, 2012

outline of the talk:

- Statistical machine translation

outline of the talk:

- Statistical machine translation
- Modeling with wta and wtt

outline of the talk:

- Statistical machine translation
- Modeling with wta and wtt
- Using automata theoretic results to “improve” modeling

outline of the talk:

- Statistical machine translation
- Modeling with wta and wtt
- Using automata theoretic results to "improve" modeling
- Summary

outline of the talk:

- Statistical machine translation **no survey!**
- Modeling with wta and wtt
- Using automata theoretic results to “improve” modeling
- Summary

given:

- source language SL
- target language TL

find:

translation $h : \mathrm{SL} \to \mathrm{TL}$

e.g.

SL = English
TL = German

s = I saw the man with the telescope
h(s) = Ich sah den Mann durch das Tel.

given:

- source language SL
- target language TL

find:

machine translation $h : \mathrm{SL} \to \mathrm{TL}$

e.g.

SL = English — s = I saw the man with the telescope

TL = German — h(s) = Ich sah den Mann durch das Tel.

given:

- source language SL
- target language TL

find:

machine translation $h : \mathrm{SL} \to \mathrm{TL}$

e.g.

SL = English   s = I saw the man with the telescope

TL = German h(s) = Ich sah den Mann durch das Tel.

machine translation $\rightsquigarrow$ statistical machine translation

given:

- source language SL
- target language TL

find:

machine translation $h : \mathrm{SL} \to \mathrm{TL}$

e.g.

SL = English — s = I saw the man with the telescope

TL = German — h(s) = Ich sah den Mann durch das Tel.

machine translation $\rightsquigarrow$ statistical machine translation

modeling - training - evaluation

given:

- source language SL
- target language TL

find:

machine translation $h : \mathrm{SL} \to \mathrm{TL}$

e.g.

SL = English   s = I saw the man with the telescope

TL = German h(s) = Ich sah den Mann durch das Tel.

machine translation $\rightsquigarrow$ statistical machine translation

modeling - training - evaluation

given:

- source language SL
- target language TL

find:

machine translation $h : \mathrm{SL} \to \mathrm{TL}$

e.g.

SL = English — s = I saw the man with the telescope

TL = German — h(s) = Ich sah den Mann durch das Tel.

machine translation $\rightsquigarrow$ statistical machine translation

modeling - training - evaluation

assumptions $\longrightarrow$ [modeling] $\longrightarrow$ $\mathcal{H}$ hypothesis space

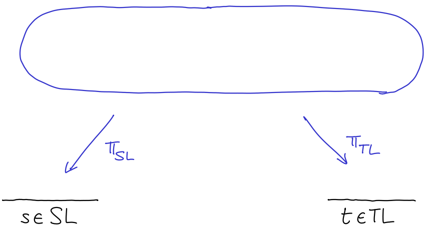
assumptions: mental work, experience, no data

hypothesis space: $\mathcal{H} \subseteq \{h \mid h : \mathrm{SL} \to \mathrm{TL}\}$

log-linear modeling:

$\overline{\hspace{3em}}_{s \in SL}$ $\overline{\hspace{3em}}_{t \in TL}$

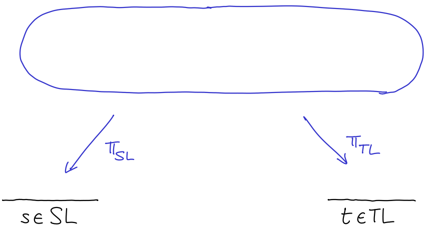
log-linear modeling:



set Y of correspondence structures [Liang et al. 06]

$\pi_{SL} : Y \to \mathrm{SL}$
$\pi_{TL} : Y \to \mathrm{TL}$

log-linear modeling:



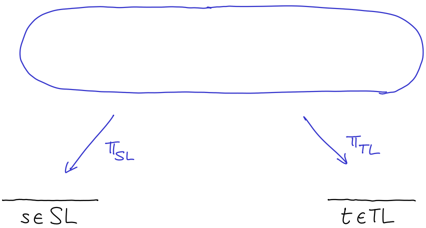
set Y of correspondence structures [Liang et al. 06]

$\pi_{SL} : Y \to \mathrm{SL}$
$\pi_{TL} : Y \to \mathrm{TL}$

hypothesis space: $\mathcal{H} = \{h_{\lambda,\Phi} \mid \lambda \in \mathbb{R}^m_{\geq 0},\ \Phi : Y \to \mathbb{R}^m\}$

$$h_{\lambda,\Phi} : \quad \mathrm{SL} \to \mathrm{TL}$$
$$s \mapsto \pi_{\mathrm{TL}}\left(\operatorname{argmax}_{\substack{y \in Y: \\ \pi_{\mathrm{SL}}(y) = s}} \lambda \cdot \Phi(y)\right)$$

log-linear modeling:



set Y of correspondence structures [Liang et al. 06]

$\pi_{SL} : Y \to \mathrm{SL}$
$\pi_{TL} : Y \to \mathrm{TL}$

hypothesis space: $\mathcal{H} = \{h_{\lambda,\Phi} \mid \lambda \in \mathbb{R}^m_{\geq 0},\ \Phi : Y \to \mathbb{R}^m\}$

$$h_{\lambda,\Phi} : \mathrm{SL} \to \mathrm{TL}$$

$$s \mapsto \pi_{\mathrm{TL}}\left(\operatorname{argmax}_{\substack{y \in Y: \\ \pi_{\mathrm{SL}}(y) = s}} \lambda \cdot \Phi(y)\right)$$

$$\lambda_1 \cdot \Phi(y)_1 + \ldots + \lambda_m \cdot \Phi(y)_m$$

hypothesis space: $\mathcal{H} = \{h_{\lambda,\Phi} \mid \lambda \in \mathbb{R}^m_{\geq 0},\ \Phi : Y \to \mathbb{R}^m\}$

$$h_{\lambda,\Phi} : \quad \mathrm{SL} \to \mathrm{TL}$$

$$s \mapsto \pi_{\mathrm{TL}} \left(\operatorname{argmax}_{\substack{y \in Y: \\ \pi_{\mathrm{SL}}(y) = s}} \lambda \cdot \Phi(y) \right)$$

$$\lambda_1 \cdot \Phi(y)_1 + \ldots + \lambda_m \cdot \Phi(y)_m$$

hypothesis space: $\mathcal{H} = \{h_{\lambda,\Phi} \mid \lambda \in \mathbb{R}^m_{\geq 0},\ \Phi : Y \to \mathbb{R}^m\}$

$$h_{\lambda,\Phi} : \quad \mathrm{SL} \to \mathrm{TL}$$
$$s \mapsto \pi_{\mathrm{TL}}\left(\operatorname{argmax}_{\substack{y \in Y: \\ \pi_{\mathrm{SL}}(y)=s}} \lambda \cdot \Phi(y)\right)$$

$$\lambda_1 \cdot \Phi(y)_1 + \ldots + \lambda_m \cdot \Phi(y)_m$$

here:

$$m = 2$$
$$\Phi(y) = (h_{\mathrm{TM}}(y), h_{\mathrm{LM}}(y))$$

h_{TM}: translation model
h_{LM}: language model

outline of the talk:

- Statistical machine translation
- Modeling with wta and wtt
- Using automata theoretic results to “improve” modeling
- Summary

first assumption:

SL and TL are the yields of
weighted recognizable tree languages

first assumption:

SL and TL are the yields of
weighted recognizable tree languages

weighted tree language: $L : T_\Sigma \to \mathbb{R}$

first assumption:

SL and TL are the yields of
weighted recognizable tree languages

weighted tree language: $L : T_{\Sigma} \to \mathbb{R}$

L is recognizable:

if there is a wta $\mathcal{A}$
which "recognizes" (computes) L

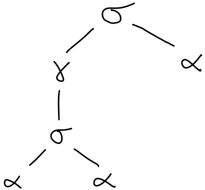
weighted tree automaton (wta) $\mathcal{A} = (Q, \Sigma, \delta, F)$

- Q finite set (states)
- Σ ranked alphabet (input symbols)
- $\delta = (\delta_\sigma \mid \sigma \in \Sigma)$ $\quad \delta_\sigma : Q^k \times Q \to \mathbb{R}$
  $\delta_\sigma(q_1 \cdots q_k, q) \in \mathbb{R}$
- $F \subseteq Q$ (final states)

weighted tree automaton (wta) $\mathcal{A} = (Q, \Sigma, \delta, F)$

- Q finite set (states)
- Σ ranked alphabet (input symbols)
- $\delta = (\delta_\sigma \mid \sigma \in \Sigma)$ $\quad \delta_\sigma : Q^k \times Q \to \mathbb{R}$
  $\delta_\sigma(q_1 \cdots q_k, q) \in \mathbb{R}$
- $F \subseteq Q$ (final states)

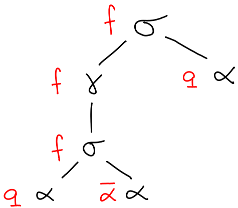
run on $\xi \in T_\Sigma$: $\quad r : \mathrm{pos}(\xi) \to Q$ $\qquad$ set of runs on ξ: $\quad R_{\mathcal{A}}(\xi)$



weighted tree automaton (wta) $\mathcal{A} = (Q, \Sigma, \delta, F)$

- Q finite set (states)
- Σ ranked alphabet (input symbols)
- $\delta = (\delta_\sigma \mid \sigma \in \Sigma)$ $\quad \delta_\sigma : Q^k \times Q \to \mathbb{R}$
  $\delta_\sigma(q_1 \cdots q_k, q) \in \mathbb{R}$
- $F \subseteq Q$ (final states)

run on $\xi \in T_\Sigma$: $r : \mathrm{pos}(\xi) \to Q$ set of runs on ξ: $R_\mathcal{A}(\xi)$

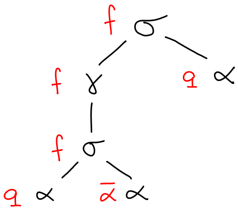


weighted tree automaton (wta) $\mathcal{A} = (Q, \Sigma, \delta, F)$

- Q finite set (states)
- Σ ranked alphabet (input symbols)
- $\delta = (\delta_\sigma \mid \sigma \in \Sigma)$ $\quad \delta_\sigma : Q^k \times Q \to \mathbb{R}$
  $\delta_\sigma(q_1 \cdots q_k, q) \in \mathbb{R}$
- $F \subseteq Q$ (final states)

run on $\xi \in T_\Sigma$: $\quad r : \mathrm{pos}(\xi) \to Q$ $\qquad$ set of runs on ξ: $\quad R_{\mathcal{A}}(\xi)$

weight of r: $\quad \mathrm{wt}(r) = \prod_{w \in \mathrm{pos}(\xi)} \delta_\sigma(r(w1) \cdots r(wk), r(w))$



σ: label of ξ at w

k: rank of σ

weighted tree automaton (wta) $\mathcal{A} = (Q, \Sigma, \delta, F)$

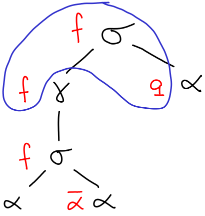
- Q finite set (states)
- Σ ranked alphabet (input symbols)
- $\delta = (\delta_\sigma \mid \sigma \in \Sigma)$ $\quad \delta_\sigma : Q^k \times Q \to \mathbb{R}$
  $\delta_\sigma(q_1 \cdots q_k, q) \in \mathbb{R}$
- $F \subseteq Q$ (final states)

run on $\xi \in T_\Sigma$: $r : \mathrm{pos}(\xi) \to Q$ set of runs on ξ: $R_\mathcal{A}(\xi)$

weight of r: $\mathrm{wt}(r) = \prod_{w \in \mathrm{pos}(\xi)} \delta_\sigma(r(w1) \cdots r(wk), r(w))$

$w = \varepsilon$

$\delta_\sigma(fq, f)$



σ: label of ξ at w

k: rank of σ

weighted tree automaton (wta) $\mathcal{A} = (Q, \Sigma, \delta, F)$

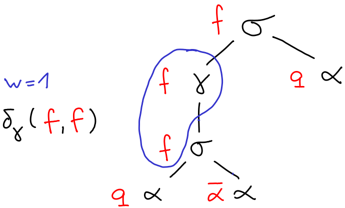
- Q finite set (states)
- Σ ranked alphabet (input symbols)
- $\delta = (\delta_\sigma \mid \sigma \in \Sigma)$ $\quad \delta_\sigma : Q^k \times Q \to \mathbb{R}$
  $\delta_\sigma(q_1 \cdots q_k, q) \in \mathbb{R}$
- $F \subseteq Q$ (final states)

run on $\xi \in T_\Sigma$: $\quad r : \mathrm{pos}(\xi) \to Q$ $\qquad$ set of runs on ξ: $\quad R_\mathcal{A}(\xi)$

weight of r: $\quad \mathrm{wt}(r) = \prod_{w \in \mathrm{pos}(\xi)} \delta_\sigma(r(w1) \cdots r(wk), r(w))$

σ: label of ξ at w

k: rank of σ



weighted tree automaton (wta) $\mathcal{A} = (Q, \Sigma, \delta, F)$

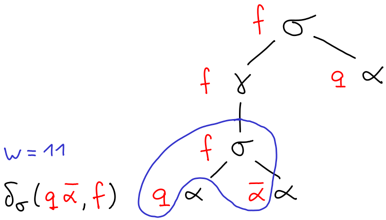
- Q finite set (states)
- Σ ranked alphabet (input symbols)
- $\delta = (\delta_\sigma \mid \sigma \in \Sigma)$ $\delta_\sigma : Q^k \times Q \to \mathbb{R}$
  $\delta_\sigma(q_1 \cdots q_k, q) \in \mathbb{R}$
- $F \subseteq Q$ (final states)

run on $\xi \in T_\Sigma$: $r : \mathrm{pos}(\xi) \to Q$ set of runs on ξ: $R_{\mathcal{A}}(\xi)$

weight of r: $\mathrm{wt}(r) = \prod_{w \in \mathrm{pos}(\xi)} \delta_\sigma(r(w1) \cdots r(wk), r(w))$

σ: label of ξ at w

k: rank of σ



weighted tree automaton (wta) $\mathcal{A} = (Q, \Sigma, \delta, F)$

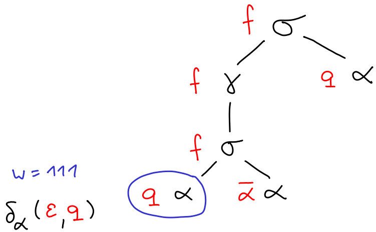
- Q finite set (states)
- Σ ranked alphabet (input symbols)
- $\delta = (\delta_\sigma \mid \sigma \in \Sigma)$ $\quad \delta_\sigma : Q^k \times Q \to \mathbb{R}$
  $\delta_\sigma(q_1 \cdots q_k, q) \in \mathbb{R}$
- $F \subseteq Q$ (final states)

run on $\xi \in T_\Sigma$: $\quad r : \mathrm{pos}(\xi) \to Q$ $\qquad$ set of runs on ξ: $\quad R_{\mathcal{A}}(\xi)$

weight of r: $\quad \mathrm{wt}(r) = \prod_{w \in \mathrm{pos}(\xi)} \delta_\sigma(r(w1) \cdots r(wk), r(w))$

σ: label of ξ at w

k: rank of σ



weighted tree automaton (wta) $\mathcal{A} = (Q, \Sigma, \delta, F)$

- Q finite set (states)
- Σ ranked alphabet (input symbols)
- $\delta = (\delta_\sigma \mid \sigma \in \Sigma)$ $\delta_\sigma : Q^k \times Q \to \mathbb{R}$
  $\delta_\sigma(q_1 \cdots q_k, q) \in \mathbb{R}$
- $F \subseteq Q$ (final states)

run on $\xi \in T_\Sigma$: $r : \mathrm{pos}(\xi) \to Q$ set of runs on ξ: $R_\mathcal{A}(\xi)$

weight of r: $\mathrm{wt}(r) = \prod_{w \in \mathrm{pos}(\xi)} \delta_\sigma(r(w1) \cdots r(wk), r(w))$

σ: label of ξ at w

k: rank of σ

weighted tree automaton (wta) $\mathcal{A} = (Q, \Sigma, \delta, F)$

- Q finite set (states)
- Σ ranked alphabet (input symbols)
- $\delta = (\delta_\sigma \mid \sigma \in \Sigma)$ $\quad \delta_\sigma : Q^k \times Q \to \mathbb{R}$
  $\delta_\sigma(q_1 \cdots q_k, q) \in \mathbb{R}$
- $F \subseteq Q$ (final states)

run on $\xi \in T_\Sigma$: $\quad r : \mathrm{pos}(\xi) \to Q$ $\qquad$ set of runs on ξ: $\quad R_{\mathcal{A}}(\xi)$

weight of r: $\quad \mathrm{wt}(r) = \prod_{w \in \mathrm{pos}(\xi)} \delta_\sigma(r(w1) \cdots r(wk), r(w))$

σ: label of ξ at w

k: rank of σ

weighted tree language recognized by $\mathcal{A}$:

$$L_{\mathcal{A}} : T_\Sigma \to \mathbb{R}, \quad L_{\mathcal{A}}(\xi) = \max_{r \in R_{\mathcal{A}}(\xi)} \mathrm{wt}(r)$$

second assumption:

translation from SL and TL is specified by a weighted tree transducer

[Yamada, Knight 01] translation from English to Japanese

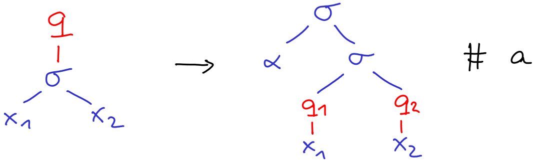
weighted tree transducer (wtt) $\mathcal{M} = (Q, \Sigma, q_0, R)$

- Q, Σ as for wta

  Σ: input <u>and</u> output symbols
- $q_0 \in Q$ (initial state)

weighted tree transducer (wtt) $\mathcal{M} = (Q, \Sigma, q_0, R)$

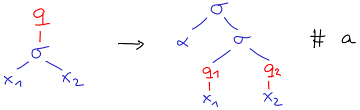
- Q, Σ as for wta — Σ: input <u>and</u> output symbols
- $q_0 \in Q$ (initial state)
- R finite set of particular term rewrite rules with weights



linear, nondeleting in $x_1, \ldots, x_k$

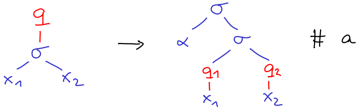
weighted tree transducer (wtt) $\mathcal{M} = (Q, \Sigma, q_0, R)$

- Q, Σ as for wta — Σ: input <u>and</u> output symbols
- $q_0 \in Q$ (initial state)
- R finite set of particular term rewrite rules with weights



weighted tree transducer (wtt) $\mathcal{M} = (Q, \Sigma, q_0, R)$

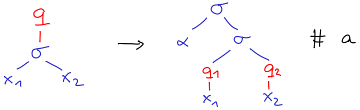
- Q, Σ as for wta — Σ: input and output symbols
- $q_0 \in Q$ (initial state)
- R finite set of particular term rewrite rules with weights



(leftmost) derivation: $d = \rho_1 \cdots \rho_n$

weighted tree transducer (wtt) $\mathcal{M} = (Q, \Sigma, q_0, R)$

- Q, Σ as for wta — Σ: input <u>and</u> output symbols
- $q_0 \in Q$ (initial state)
- R finite set of particular term rewrite rules with weights

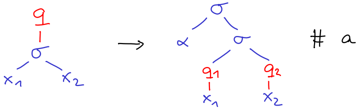


(leftmost) derivation: $d = \rho_1 \cdots \rho_n$

weight of a derivation d: $\mathrm{wt}(d) = \prod_{i=1}^{n} \mathrm{wt}(\rho_i)$

weighted tree transducer (wtt) $\mathcal{M} = (Q, \Sigma, q_0, R)$

- Q, Σ as for wta — Σ: input <u>and</u> output symbols
- $q_0 \in Q$ (initial state)
- R finite set of particular term rewrite rules with weights



(leftmost) derivation: $d = \rho_1 \cdots \rho_n$

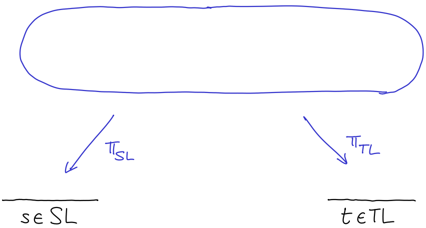
weight of a derivation d: $\mathrm{wt}(d) = \prod_{i=1}^{n} \mathrm{wt}(\rho_i)$

weighted tree transformation computed by $\mathcal{M}$:

$$\tau_{\mathcal{M}} : T_\Sigma \times T_\Sigma \to \mathbb{R}, \quad \tau_{\mathcal{M}}(\xi_1, \xi_2) = \max_{\substack{d \in D_{\mathcal{M}}: \\ \pi(d) = (\xi_1, \xi_2)}} \mathrm{wt}(d)$$

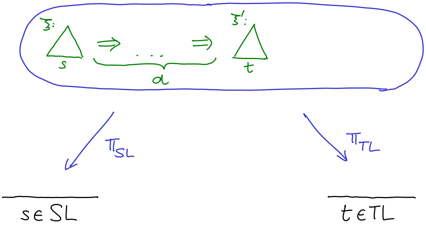
$D_{\mathcal{M}}$: set of all derivations

log-linear modeling with wtt and wta:



set Y of correspondence structures [Liang et al. 06]

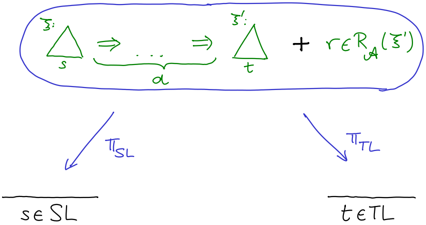
log-linear modeling with wtt and wta:



set Y of correspondence structures [Liang et al. 06]

wtt $\mathcal{M}$ as translation model; $d \in D_{\mathcal{M}}$

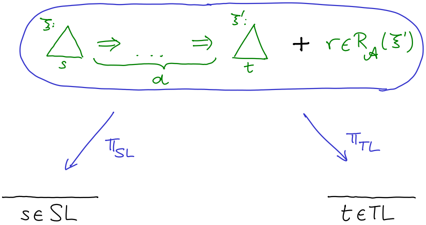
log-linear modeling with wtt and wta:



set Y of correspondence structures [Liang et al. 06]

wtt $\mathcal{M}$ as translation model; $d \in D_{\mathcal{M}}$
wta $\mathcal{A}$ as language model; $r \in R_{\mathcal{A}}$

log-linear modeling with wtt and wta:



set Y of correspondence structures [Liang et al. 06]

wtt $\mathcal{M}$ as translation model; $d \in D_{\mathcal{M}}$
wta $\mathcal{A}$ as language model; $r \in R_{\mathcal{A}}$

$$Y = \{(d, r) \in D_{\mathcal{M}} \times R_{\mathcal{A}} \mid r \in R_{\mathcal{A}}(\mathrm{last}(d))\}$$

$$\pi_{SL}(d, r) = \mathrm{yield}(\mathrm{first}(d))$$
$$\pi_{TL}(d, r) = \mathrm{yield}(\mathrm{last}(d))$$

wtt $\mathcal{M}$ as translation model; $d \in D_{\mathcal{M}}$
wta $\mathcal{A}$ as language model; $r \in R_{\mathcal{A}}$

$$Y = \{(d, r) \in D_{\mathcal{M}} \times R_{\mathcal{A}} \mid r \in R_{\mathcal{A}}(\mathrm{last}(d))\}$$

wtt $\mathcal{M}$ as translation model; $d \in D_{\mathcal{M}}$
wta $\mathcal{A}$ as language model; $r \in R_{\mathcal{A}}$

$$Y = \{(d, r) \in D_{\mathcal{M}} \times R_{\mathcal{A}} \mid r \in R_{\mathcal{A}}(\mathrm{last}(d))\}$$

here:

$$\begin{aligned} m &= 2 \\ \Phi(d, r) &= \Big(\log \mathrm{wt}_{\mathcal{M}}(d),\ \log \mathrm{wt}_{\mathcal{A}}(r)\Big) \end{aligned}$$

wtt $\mathcal{M}$ as translation model; $d \in D_{\mathcal{M}}$
wta $\mathcal{A}$ as language model; $r \in R_{\mathcal{A}}$

$$Y = \{(d, r) \in D_{\mathcal{M}} \times R_{\mathcal{A}} \mid r \in R_{\mathcal{A}}(\mathrm{last}(d))\}$$

here:

$$\begin{aligned} m &= 2 \\ \Phi(d, r) &= \Big(\log \mathrm{wt}_{\mathcal{M}}(d),\ \log \mathrm{wt}_{\mathcal{A}}(r)\Big) \end{aligned}$$

$$\mathcal{H} = \{h_{\lambda,\mathcal{M},\mathcal{A}} \mid \lambda \in \mathbb{R}^2_{\geq 0}, \quad \text{wtt } \mathcal{M}, \quad \text{wta } \mathcal{A}\}$$

$$\begin{aligned} h_{\lambda,\mathcal{M},\mathcal{A}} : \quad & \mathrm{SL} \to \mathrm{TL} \\ & s \mapsto \pi_{\mathrm{TL}}\left(\operatorname*{argmax}_{\substack{(d,r)\in Y: \\ \pi_{\mathrm{SL}}(d,r)=s}} \mathrm{wt}_{\mathcal{M}}(d)^{\lambda_1} \cdot \mathrm{wt}_{\mathcal{A}}(r)^{\lambda_2}\right) \end{aligned}$$

outline of the talk:

- Statistical machine translation
- Modeling with wta and wtt
- Using automata theoretic results to “improve” modeling
  - Weight exponentiation
  - Output product
- Summary

$$
\begin{aligned}
h_{\lambda,\mathcal{M},\mathcal{A}} : \quad \mathrm{SL} &\to \mathrm{TL} \\
s &\mapsto \pi_{\mathrm{TL}}\left(\operatorname{argmax}_{\substack{(d,r)\in Y: \\ \pi_{\mathrm{SL}}(d,r)=s}} \mathrm{wt}_{\mathcal{M}}(d)^{\lambda_1} \cdot \mathrm{wt}_{\mathcal{A}}(r)^{\lambda_2} \right)
\end{aligned}
$$

$$
\begin{aligned}
h_{\lambda,\mathcal{M},\mathcal{A}} : \quad \mathrm{SL} &\rightarrow \mathrm{TL} \\
s &\mapsto \pi_{\mathrm{TL}}\left(\operatorname{argmax}_{\substack{(d,r)\in Y: \\ \pi_{\mathrm{SL}}(d,r)=s}} \mathrm{wt}_{\mathcal{M}}(d)^{\lambda_1}\cdot \mathrm{wt}_{\mathcal{A}}(r)^{\lambda_2}\right) \\
&= \pi_{\mathrm{TL}}\left(\operatorname{argmax}_{\substack{(d,r)\in Y: \\ \pi_{\mathrm{SL}}(d,r)=s}} \mathrm{wt}_{\mathcal{M}}(d)\cdot \mathrm{wt}_{\mathcal{A}}(r)^{\lambda_2/\lambda_1}\right)
\end{aligned}
$$

$$\begin{aligned} h_{\lambda,\mathcal{M},\mathcal{A}}: \quad \mathrm{SL} &\to \mathrm{TL} \\ s &\mapsto \pi_{\mathrm{TL}}\left(\operatorname{argmax}_{\substack{(d,r)\in Y:\\ \pi_{\mathrm{SL}}(d,r)=s}} \mathrm{wt}_{\mathcal{M}}(d)^{\lambda_1} \cdot \mathrm{wt}_{\mathcal{A}}(r)^{\lambda_2}\right) \\ &= \pi_{\mathrm{TL}}\left(\operatorname{argmax}_{\substack{(d,r)\in Y:\\ \pi_{\mathrm{SL}}(d,r)=s}} \mathrm{wt}_{\mathcal{M}}(d) \cdot \mathrm{wt}_{\mathcal{A}}(r)^{\lambda_2/\lambda_1}\right) \end{aligned}$$

Lemma: Let $\mathcal{A}$ be a wta and $\lambda \in \mathbb{R}_{\geq 0}$.

$$
\begin{aligned}
h_{\lambda,\mathcal{M},\mathcal{A}}: \quad \mathrm{SL} &\to \mathrm{TL} \\
s &\mapsto \pi_{\mathrm{TL}}\left(\operatorname*{argmax}_{\substack{(d,r)\in Y:\\ \pi_{\mathrm{SL}}(d,r)=s}} \mathrm{wt}_{\mathcal{M}}(d)^{\lambda_1}\cdot \mathrm{wt}_{\mathcal{A}}(r)^{\lambda_2}\right) \\
&= \pi_{\mathrm{TL}}\left(\operatorname*{argmax}_{\substack{(d,r)\in Y:\\ \pi_{\mathrm{SL}}(d,r)=s}} \mathrm{wt}_{\mathcal{M}}(d)\cdot \mathrm{wt}_{\mathcal{A}}(r)^{\lambda_2/\lambda_1}\right)
\end{aligned}
$$

Lemma: Let $\mathcal{A}$ be a wta and $\lambda \in \mathbb{R}_{\geq 0}$.
There is a wta $\mathcal{A}'$ s.t. $Q_{\mathcal{A}'} = Q_{\mathcal{A}}$ and
$\mathrm{wt}_{\mathcal{A}'}(r) = \mathrm{wt}_{\mathcal{A}}(r)^{\lambda}$ for every r.

$$h_{\lambda,\mathcal{M},\mathcal{A}}: \quad \mathrm{SL} \rightarrow \mathrm{TL}$$

$$s \mapsto \pi_{\mathrm{TL}}\left(\operatorname{argmax}_{\substack{(d,r)\in Y:\\ \pi_{\mathrm{SL}}(d,r)=s}} \mathrm{wt}_{\mathcal{M}}(d)^{\lambda_1} \cdot \mathrm{wt}_{\mathcal{A}}(r)^{\lambda_2}\right)$$

$$= \pi_{\mathrm{TL}}\left(\operatorname{argmax}_{\substack{(d,r)\in Y:\\ \pi_{\mathrm{SL}}(d,r)=s}} \mathrm{wt}_{\mathcal{M}}(d) \cdot \mathrm{wt}_{\mathcal{A}}(r)^{\lambda_2/\lambda_1}\right)$$

Lemma: Let $\mathcal{A}$ be a wta and $\lambda \in \mathbb{R}_{\geq 0}$.
There is a wta $\mathcal{A}'$ s.t. $Q_{\mathcal{A}'} = Q_{\mathcal{A}}$ and
$\mathrm{wt}_{\mathcal{A}'}(r) = \mathrm{wt}_{\mathcal{A}}(r)^{\lambda}$ for every r.

$$= \pi_{\mathrm{TL}}\left(\operatorname{argmax}_{\substack{(d,r)\in Y':\\ \pi_{\mathrm{SL}}(d,r)=s}} \mathrm{wt}_{\mathcal{M}}(d) \cdot \mathrm{wt}_{\mathcal{A}'}(r)\right)$$

outline of the talk:

- Statistical machine translation
- Modeling with wta and wtt
- Using automata theoretic results to “improve” modeling
  - Weight exponentiation
  - Output product
- Summary

Let $\tau : T_\Sigma \times T_\Sigma \to \mathbb{R}$ and $L : T_\Sigma \to \mathbb{R}$

Let $\tau : T_\Sigma \times T_\Sigma \to \mathbb{R}$ and $L : T_\Sigma \to \mathbb{R}$

output product of τ and L:

$$\tau \triangleright L : T_\Sigma \times T_\Sigma \to \mathbb{R}$$

$$(\tau \triangleright L)(\xi, \zeta) \mapsto \tau(\xi, \zeta) \cdot L(\zeta)$$

Let $\tau : T_\Sigma \times T_\Sigma \to \mathbb{R}$ and $L : T_\Sigma \to \mathbb{R}$

output product of τ and L:

$$\tau \triangleright L : T_\Sigma \times T_\Sigma \to \mathbb{R}$$

$$(\tau \triangleright L)(\xi, \zeta) \mapsto \tau(\xi, \zeta) \cdot L(\zeta)$$

input product of L and τ:

$$L \triangleleft \tau : T_\Sigma \times T_\Sigma \to \mathbb{R}$$

$$(L \triangleleft \tau)(\xi, \zeta) \mapsto L(\xi) \cdot \tau(\xi, \zeta)$$

Theorem [Maletti 06]: Let $\mathcal{M}$ wtt and $\mathcal{A}$ wta.

Theorem [Maletti 06]: Let $\mathcal{M}$ wtt and $\mathcal{A}$ wta.
There is a wtt $\mathcal{M} \triangleright \mathcal{A}$ such that: $\tau_{\mathcal{M} \triangleright \mathcal{A}} = \tau_{\mathcal{M}} \triangleright L_{\mathcal{A}}$

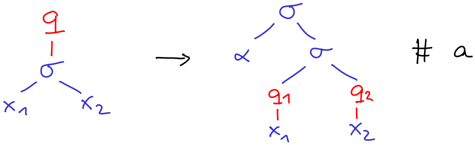
Theorem [Maletti 06]: Let $\mathcal{M}$ wtt and $\mathcal{A}$ wta.
There is a wtt $\mathcal{M} \triangleright \mathcal{A}$ such that: $\tau_{\mathcal{M} \triangleright \mathcal{A}} = \tau_{\mathcal{M}} \triangleright L_{\mathcal{A}}$

Proof: [Baker 79, Engelfriet, Fülöp,V. 02]

Theorem [Maletti 06]: Let $\mathcal{M}$ wtt and $\mathcal{A}$ wta.
There is a wtt $\mathcal{M} \triangleright \mathcal{A}$ such that: $\tau_{\mathcal{M} \triangleright \mathcal{A}} = \tau_{\mathcal{M}} \triangleright L_{\mathcal{A}}$

Proof: [Baker 79, Engelfriet, Fülöp,V. 02]

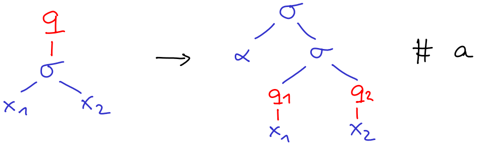
rule of $\mathcal{M}$:



Theorem [Maletti 06]: Let $\mathcal{M}$ wtt and $\mathcal{A}$ wta.
There is a wtt $\mathcal{M} \triangleright \mathcal{A}$ such that: $\tau_{\mathcal{M} \triangleright \mathcal{A}} = \tau_{\mathcal{M}} \triangleright L_{\mathcal{A}}$

Proof: [Baker 79, Engelfriet, Fülöp,V. 02]

rule of $\mathcal{M}$:

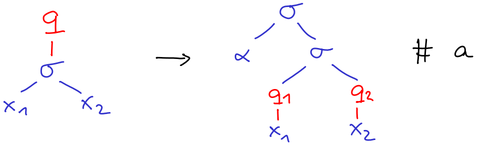


states of $\mathcal{A}$: p, p_1, p_2

Theorem [Maletti 06]: Let $\mathcal{M}$ wtt and $\mathcal{A}$ wta.
There is a wtt $\mathcal{M} \triangleright \mathcal{A}$ such that: $\tau_{\mathcal{M} \triangleright \mathcal{A}} = \tau_{\mathcal{M}} \triangleright L_{\mathcal{A}}$

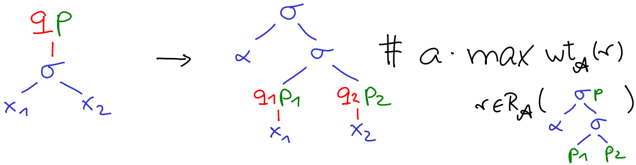
Proof: [Baker 79, Engelfriet, Fülöp,V. 02]

rule of $\mathcal{M}$:

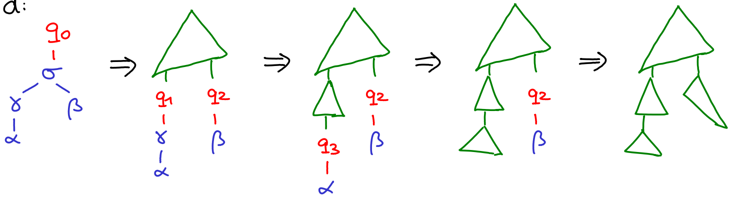


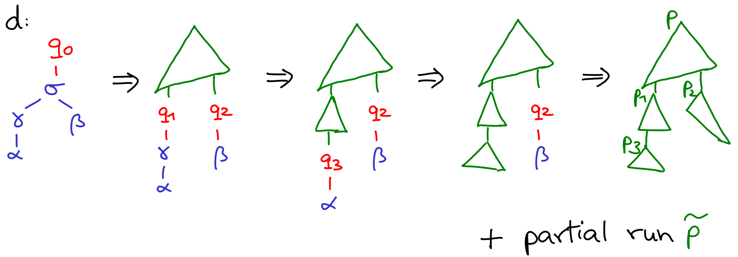
states of $\mathcal{A}$: p, p_1, p_2

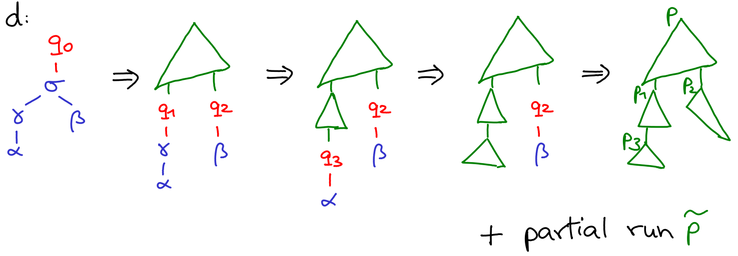
rule of $\mathcal{M} \triangleright \mathcal{A}$:

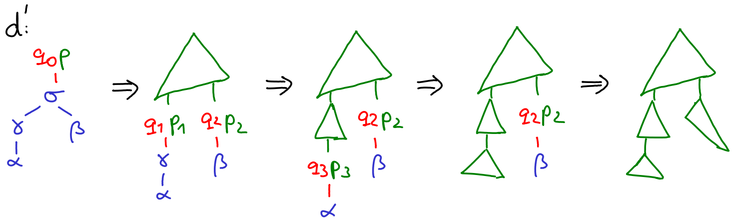


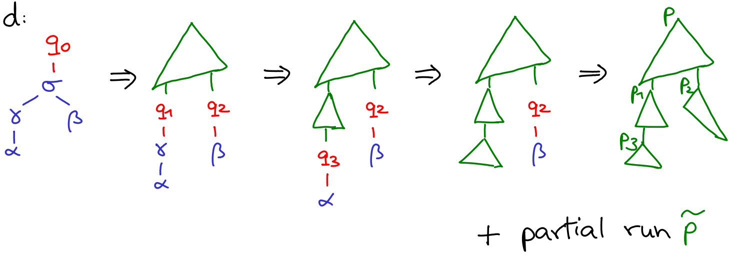
d:

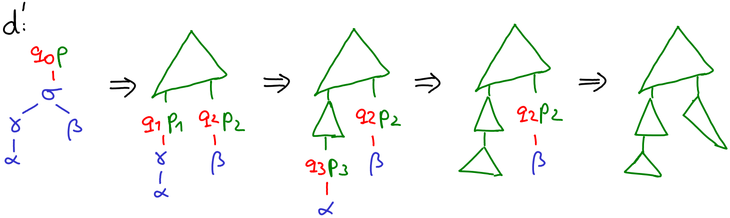



d:
q0
σ
γ
β
α
q1
q2
q3
p
p1
p2
p3
+ partial run p̃


d:
q0
σ
γ
β
α
⇒
q1
q2
q3
p
p1
p2
p3
+ partial run p̃


d':
q0 p
σ
γ
β
α
⇒
q1 p1
q2 p2
q3 p3





$\varphi : D_{\mathcal{M}'} \rightarrow \{(d, \tilde{p}) \mid d \in D_{\mathcal{M}}, \tilde{p} \in R_{\mathcal{A}}^{\text{partial}}(d)\}$ bijection

$\mathrm{wt}(d') = \mathrm{wt}(d) \cdot \max_{r \in \mathrm{completion}(\tilde{p})} \mathrm{wt}(r)$

recall:

$$h_{\lambda,\mathcal{M},\mathcal{A}} : \mathrm{SL} \to \mathrm{TL}$$

$$s \quad \mapsto \quad \pi_{\mathrm{TL}}\left(\operatorname{argmax}_{\substack{(d,r)\in Y': \\ \pi_{\mathrm{SL}}(d,r)=s}} \mathrm{wt}_{\mathcal{M}}(d) \cdot \mathrm{wt}_{\mathcal{A}'}(r)\right)$$

$h_{\lambda,\mathcal{M},\mathcal{A}} : \mathrm{SL} \to \mathrm{TL}$

$$s \mapsto \pi_{\mathrm{TL}}\left(\operatorname{argmax}_{\substack{(d,r)\in Y':\\ \pi_{\mathrm{SL}}(d,r)=s}} \mathrm{wt}_{\mathcal{M}}(d)\cdot \mathrm{wt}_{\mathcal{A}'}(r)\right)$$

$$= \pi_{\mathrm{TL}}\left(\operatorname{argmax}_{\substack{d\in D_{\mathcal{M}}\\ \pi_{\mathrm{SL}}(d)=s}} \mathrm{wt}_{\mathcal{M}}(d)\cdot \max_{r\in R_{\mathcal{A}'}(\mathrm{last}(d))} \mathrm{wt}_{\mathcal{A}'}(r)\right)$$

$$h_{\lambda,\mathcal{M},\mathcal{A}} : \mathrm{SL} \to \mathrm{TL}$$

$$
\begin{aligned}
s &\mapsto \pi_{\mathrm{TL}}\left(\operatorname{argmax}_{\substack{(d,r)\in Y':\\ \pi_{\mathrm{SL}}(d,r)=s}} \mathrm{wt}_{\mathcal{M}}(d) \cdot \mathrm{wt}_{\mathcal{A}'}(r)\right) \\
&= \pi_{\mathrm{TL}}\left(\operatorname{argmax}_{\substack{d\in D_{\mathcal{M}}\\ \pi_{\mathrm{SL}}(d)=s}} \mathrm{wt}_{\mathcal{M}}(d) \cdot \max_{r\in R_{\mathcal{A}'}(\mathrm{last}(d))} \mathrm{wt}_{\mathcal{A}'}(r)\right) \\
&= \pi_{\mathrm{TL}}\left(\operatorname{argmax}_{\substack{d\in D_{\mathcal{M}}\\ \pi_{\mathrm{SL}}(d)=s}} \mathrm{wt}_{\mathcal{M}}(d) \cdot \max_{\tilde{p}\in R^{\mathrm{p}}_{\mathcal{A}'}(d)} \max_{r\in \mathrm{compl.}(\tilde{p})} \mathrm{wt}_{\mathcal{A}'}(r)\right)
\end{aligned}
$$

$h_{\lambda,\mathcal{M},\mathcal{A}} : \mathrm{SL} \to \mathrm{TL}$

$$
\begin{aligned}
s \;\mapsto\;& \pi_{\mathrm{TL}}\left(\operatorname{argmax}_{\substack{(d,r)\in Y':\\ \pi_{\mathrm{SL}}(d,r)=s}} \mathrm{wt}_{\mathcal{M}}(d)\cdot \mathrm{wt}_{\mathcal{A}'}(r)\right)\\
=\;& \pi_{\mathrm{TL}}\left(\operatorname{argmax}_{\substack{d\in D_{\mathcal{M}}\\ \pi_{\mathrm{SL}}(d)=s}} \mathrm{wt}_{\mathcal{M}}(d)\cdot \max_{r\in R_{\mathcal{A}'}(\mathrm{last}(d))} \mathrm{wt}_{\mathcal{A}'}(r)\right)\\
=\;& \pi_{\mathrm{TL}}\left(\operatorname{argmax}_{\substack{d\in D_{\mathcal{M}}\\ \pi_{\mathrm{SL}}(d)=s}} \mathrm{wt}_{\mathcal{M}}(d)\cdot \max_{\tilde{p}\in R^{\mathrm{p}}_{\mathcal{A}'}(d)} \max_{r\in \mathrm{compl.}(\tilde{p})} \mathrm{wt}_{\mathcal{A}'}(r)\right)\\
=\;& \pi_{\mathrm{TL}}\left(\operatorname{argmax}_{\substack{d\in D_{\mathcal{M}}\\ \pi_{\mathrm{SL}}(d)=s\\ \tilde{p}\in R^{\mathrm{p}}_{\mathcal{A}'}(d)}} \mathrm{wt}_{\mathcal{M}}(d)\cdot \max_{r\in \mathrm{compl.}(\tilde{p})} \mathrm{wt}_{\mathcal{A}'}(r)\right)
\end{aligned}
$$

$$h_{\lambda,\mathcal{M},\mathcal{A}} : \mathrm{SL} \to \mathrm{TL}$$

$$\begin{aligned}
s \quad \mapsto \quad & \pi_{\mathrm{TL}}\left(\operatorname{argmax}_{\substack{(d,r)\in Y':\\ \pi_{\mathrm{SL}}(d,r)=s}} \mathrm{wt}_{\mathcal{M}}(d) \cdot \mathrm{wt}_{\mathcal{A}'}(r)\right) \\
= \quad & \pi_{\mathrm{TL}}\left(\operatorname{argmax}_{\substack{d\in D_{\mathcal{M}}\\ \pi_{\mathrm{SL}}(d)=s}} \mathrm{wt}_{\mathcal{M}}(d) \cdot \max_{r\in R_{\mathcal{A}'}(\mathrm{last}(d))} \mathrm{wt}_{\mathcal{A}'}(r)\right) \\
= \quad & \pi_{\mathrm{TL}}\left(\operatorname{argmax}_{\substack{d\in D_{\mathcal{M}}\\ \pi_{\mathrm{SL}}(d)=s}} \mathrm{wt}_{\mathcal{M}}(d) \cdot \max_{\tilde{p}\in R^{\mathrm{p}}_{\mathcal{A}'}(d)} \max_{r\in \mathrm{compl.}(\tilde{p})} \mathrm{wt}_{\mathcal{A}'}(r)\right) \\
= \quad & \pi_{\mathrm{TL}}\left(\operatorname{argmax}_{\substack{d\in D_{\mathcal{M}}\\ \pi_{\mathrm{SL}}(d)=s\\ \tilde{p}\in R^{\mathrm{p}}_{\mathcal{A}'}(d)}} \mathrm{wt}_{\mathcal{M}}(d) \cdot \max_{r\in \mathrm{compl.}(\tilde{p})} \mathrm{wt}_{\mathcal{A}'}(r)\right) \\
= \quad & \pi_{\mathrm{TL}}\left(\operatorname{argmax}_{\substack{d\in D_{\mathcal{M}\triangleright\mathcal{A}'}\\ \pi_{\mathrm{SL}}(d)=s}} \mathrm{wt}_{\mathcal{M}\triangleright\mathcal{A}'}(d)\right)
\end{aligned}$$

recall:

Theorem [Maletti 06]: Let $\mathcal{M}$ wtt and $\mathcal{A}$ wta.
There is a wtt $\mathcal{M} \triangleright \mathcal{A}$ such that: $\tau_{\mathcal{M} \triangleright \mathcal{A}} = \tau_{\mathcal{M}} \triangleright L_{\mathcal{A}}$

recall:

Theorem [Maletti 06]: Let $\mathcal{M}$ wtt and $\mathcal{A}$ wta.
There is a wtt $\mathcal{M} \triangleright \mathcal{A}$ such that: $\tau_{\mathcal{M} \triangleright \mathcal{A}} = \tau_{\mathcal{M}} \triangleright L_{\mathcal{A}}$

generalization to mildly context-sensitive languages

Theorem [Büchse, Nederhof, V. 11]:
Let $\mathcal{M}$ synchronized tree-adjoining grammar (STAG)
and $\mathcal{A}$ wta.
There is an STAG $\mathcal{M} \triangleright \mathcal{A}$ such that: $\tau_{\mathcal{M} \triangleright \mathcal{A}} = \tau_{\mathcal{M}} \triangleright L_{\mathcal{A}}$

recall:

Theorem [Maletti 06]: Let $\mathcal{M}$ wtt and $\mathcal{A}$ wta.
There is a wtt $\mathcal{M} \triangleright \mathcal{A}$ such that: $\tau_{\mathcal{M} \triangleright \mathcal{A}} = \tau_{\mathcal{M}} \triangleright L_{\mathcal{A}}$

generalization to mildly context-sensitive languages

Theorem [Büchse, Nederhof, V. 11]:
Let $\mathcal{M}$ synchronized tree-adjoining grammar (STAG)
and $\mathcal{A}$ wta.
There is an STAG $\mathcal{M} \triangleright \mathcal{A}$ such that: $\tau_{\mathcal{M} \triangleright \mathcal{A}} = \tau_{\mathcal{M}} \triangleright L_{\mathcal{A}}$

Theorem [Nederhof, V. 12]:
Let $\mathcal{M}$ synchronized context-free tree grammar (SCFTG)
and $\mathcal{A}$ wta.
There is an SCFTG $\mathcal{M} \triangleright \mathcal{A}$ such that: $\tau_{\mathcal{M} \triangleright \mathcal{A}} = \tau_{\mathcal{M}} \triangleright L_{\mathcal{A}}$

outline of the talk:

- Statistical machine translation
- Modeling with wta and wtt
- Using automata theoretic results to “improve” modeling
- Summary

- model for translation from SL to TL: wtt $\mathcal{M}$
- model for TL: wta $\mathcal{A}$
- sentence s of SL

- model for translation from SL to TL: wtt $\mathcal{M}$
- model for TL: wta $\mathcal{A}$
- sentence s of SL

$$h_{\lambda,\mathcal{M},\mathcal{A}}(s) \quad = \quad \pi_{\mathrm{TL}}\left(\operatorname{argmax}_{\substack{(d,r)\in Y:\\ \pi_{\mathrm{SL}}(d,r)=s}} \mathrm{wt}_{\mathcal{M}}(d)^{\lambda_1}\cdot \mathrm{wt}_{\mathcal{A}}(r)^{\lambda_2}\right)$$

- model for translation from SL to TL: wtt $\mathcal{M}$
- model for TL: wta $\mathcal{A}$
- sentence s of SL

$$h_{\lambda,\mathcal{M},\mathcal{A}}(s) = \pi_{\mathrm{TL}}\left(\operatorname{argmax}_{\substack{(d,r)\in Y: \\ \pi_{\mathrm{SL}}(d,r)=s}} \mathrm{wt}_{\mathcal{M}}(d)^{\lambda_1} \cdot \mathrm{wt}_{\mathcal{A}}(r)^{\lambda_2}\right)$$

$$\text{(weight exp.)} = \pi_{\mathrm{TL}}\left(\operatorname{argmax}_{\substack{(d,r)\in Y': \\ \pi_{\mathrm{SL}}(d,r)=s}} \mathrm{wt}_{\mathcal{M}}(d) \cdot \mathrm{wt}_{\mathcal{A}'}(r)\right)$$

- model for translation from SL to TL: wtt $\mathcal{M}$
- model for TL: wta $\mathcal{A}$
- sentence s of SL

$$
\begin{aligned}
h_{\lambda,\mathcal{M},\mathcal{A}}(s) &= \pi_{\mathrm{TL}}\left(\operatorname{argmax}_{\substack{(d,r)\in Y:\\ \pi_{\mathrm{SL}}(d,r)=s}} \mathrm{wt}_{\mathcal{M}}(d)^{\lambda_1}\cdot \mathrm{wt}_{\mathcal{A}}(r)^{\lambda_2}\right)\\
\text{(weight exp.)} &= \pi_{\mathrm{TL}}\left(\operatorname{argmax}_{\substack{(d,r)\in Y':\\ \pi_{\mathrm{SL}}(d,r)=s}} \mathrm{wt}_{\mathcal{M}}(d)\cdot \mathrm{wt}_{\mathcal{A}'}(r)\right)\\
\text{(output product)} &= \pi_{\mathrm{TL}}\left(\operatorname{argmax}_{\substack{d\in D_{\mathcal{M}\triangleright\mathcal{A}'}:\\ \pi_{\mathrm{SL}}(d)=s}} \mathrm{wt}_{\mathcal{M}\triangleright\mathcal{A}'}(d)\right)
\end{aligned}
$$

- model for translation from SL to TL: wtt $\mathcal{M}$
- model for TL: wta $\mathcal{A}$
- sentence s of SL

$$h_{\lambda,\mathcal{M},\mathcal{A}}(s) = \pi_{\mathrm{TL}}\left(\operatorname{argmax}_{\substack{(d,r)\in Y:\\ \pi_{\mathrm{SL}}(d,r)=s}} \mathrm{wt}_{\mathcal{M}}(d)^{\lambda_1}\cdot \mathrm{wt}_{\mathcal{A}}(r)^{\lambda_2}\right)$$

(weight exp.) $$= \pi_{\mathrm{TL}}\left(\operatorname{argmax}_{\substack{(d,r)\in Y':\\ \pi_{\mathrm{SL}}(d,r)=s}} \mathrm{wt}_{\mathcal{M}}(d)\cdot \mathrm{wt}_{\mathcal{A}'}(r)\right)$$

(output product) $$= \pi_{\mathrm{TL}}\left(\operatorname{argmax}_{\substack{d\in D_{\mathcal{M}\triangleright\mathcal{A}'}:\\ \pi_{\mathrm{SL}}(d)=s}} \mathrm{wt}_{\mathcal{M}\triangleright\mathcal{A}'}(d)\right)$$

(B,S,P; input product) $$= \pi_{\mathrm{TL}}\left(\operatorname{argmax}_{d\in D_{\mathcal{A}_s\triangleleft(\mathcal{M}\triangleright\mathcal{A})}} \mathrm{wt}_{\mathcal{A}_s\triangleleft(\mathcal{M}\triangleright\mathcal{A}')}(d)\right)$$

- model for translation from SL to TL: wtt $\mathcal{M}$
- model for TL: wta $\mathcal{A}$
- sentence s of SL

$$h_{\lambda,\mathcal{M},\mathcal{A}}(s) = \pi_{\mathrm{TL}}\left(\operatorname{argmax}_{\substack{(d,r)\in Y:\\ \pi_{\mathrm{SL}}(d,r)=s}} \mathrm{wt}_{\mathcal{M}}(d)^{\lambda_1}\cdot \mathrm{wt}_{\mathcal{A}}(r)^{\lambda_2}\right)$$

(weight exp.) $$= \pi_{\mathrm{TL}}\left(\operatorname{argmax}_{\substack{(d,r)\in Y':\\ \pi_{\mathrm{SL}}(d,r)=s}} \mathrm{wt}_{\mathcal{M}}(d)\cdot \mathrm{wt}_{\mathcal{A}'}(r)\right)$$

(output product) $$= \pi_{\mathrm{TL}}\left(\operatorname{argmax}_{\substack{d\in D_{\mathcal{M}\triangleright\mathcal{A}'}:\\ \pi_{\mathrm{SL}}(d)=s}} \mathrm{wt}_{\mathcal{M}\triangleright\mathcal{A}'}(d)\right)$$

(B,S,P; input product) $$= \pi_{\mathrm{TL}}\left(\operatorname{argmax}_{d\in D_{\mathcal{A}_s\triangleleft(\mathcal{M}\triangleright\mathcal{A})}} \mathrm{wt}_{\mathcal{A}_s\triangleleft(\mathcal{M}\triangleright\mathcal{A}')}(d)\right)$$

$$= \pi_{\mathrm{TL}}\big(\,\mathrm{Knuth}(\mathcal{A}_s\triangleleft(\mathcal{M}\triangleright\mathcal{A}'))\,\big)$$

impression: SMT is “playing with formulas” impression: SMT is "playing with formulas"

but: SMT is an engineering task!

impression: SMT is “playing with formulas”

but: SMT is an engineering task!

weighted tree automata and weighted tree transducers

can help in **modeling**

statistical machine translation of natural languages

References:

- [Baker 79] Composition of top-down and bottom-up tree transductions
- [Bar-Hillel, Shamir, Perles 61] On formal properties of simple phrase structure grammars
- [Büchse, Nederhof, V. 11] Tree Parsing with Synchronous Tree-Adjoining Grammars
- [Engelfriet, Fülöp, V. 02] Bottom-up and Top-down Tree Series Transformations
- [Fülöp, V. 09] Weighted tree automata and tree transducers
- [Knight et al. 03-...] ...
- [Knuth 77] A generalization of Dijkstra's algorithm
- [Lopez 08] Statistical Machine Translation
- [Liang, Bouchard-Côté, Klein, Taskar 06] An End-to-End Discriminative Approach to Machine Translation
- [Maletti 06] Compositions of Tree Series Transformations
- [Maletti, Satta 09] Parsing Algorithms based on Tree Automata
- [Nederhof, V. 12] Synchronous Context-Free Tree Grammars
- [Thatcher 67] Characterizing derivation trees of context-free grammars through a generalization of finite automata theory
- [Yamada, Knight 01] A syntax-based statistical translation model

[Knight et al. 03-...]

- [Charniak, Knight, Yamada 03] Syntax-based language models for statistical machine translation
- [Galley, Hopkins, Knight, Marcu 04] What's in a translation rule?
- [Graehl, Knight 04] Training tree transducers
- [Graehl, Knight 05] An overview of probabilistic tree transducers for natural language processing

[Knight et al. 03-...]

- [Charniak, Knight, Yamada 03] Syntax-based language models for statistical machine translation
- [Galley, Hopkins, Knight, Marcu 04] What's in a translation rule?
- [Graehl, Knight 04] Training tree transducers
- [Graehl, Knight 05] An overview of probabilistic tree transducers for natural language processing

[Maletti 11]
Survey: Weighted Extended Top-down Tree Transducers –
Part III: Applications in Machine Translation

[Knight et al. 03-...]

- [Charniak, Knight, Yamada 03] Syntax-based language models for statistical machine translation
- [Galley, Hopkins, Knight, Marcu 04] What's in a translation rule?
- [Graehl, Knight 04] Training tree transducers
- [Graehl, Knight 05] An overview of probabilistic tree transducers for natural language processing

[Maletti 11]
Survey: Weighted Extended Top-down Tree Transducers –
Part III: Applications in Machine Translation

Thank you!