Applications of Tree Automata Theory
Lecture VI: Back to Machine Translation

Andreas Maletti

Institute of Computer Science
Universität Leipzig, Germany

on leave from: Institute for Natural Language Processing
Universität Stuttgart, Germany

maletti@ims.uni-stuttgart.de

Yekaterinburg — August 25, 2014

Roadmap

1. Theory of Tree Automata
2. Parsing — Basics and Evaluation
3. Parsing — Advanced Topics
4. Machine Translation — Basics and Evaluation
5. Theory of Tree Transducers
6. Machine Translation — Advanced Topics

Always ask questions right away!

Tree Transducers in Machine Translation

Important relations

- SCFG = synchronous context-free grammar LTG-LTG
  [CHIANG, 2007] (synchronous local tree grammar)
  $\subseteq$ ln-TOP special top-down tree transducer
- STSG = synchronous tree substitution grammar TSG-TSG
  [EISNER, 2003]
  $\subseteq$ ln-XTOP special extended top-down tree transducer
- STAG = synchronous tree adjunction grammar TAG-TAG
  [SHIEBER, SCHABES, 1990]
- SCFTG = synchronous context-free tree grammar
  [NEDERHOF, VOGLER, 2012] CFTG-CFTG

Tree Transducers in Machine Translation

Towards asymetric relations

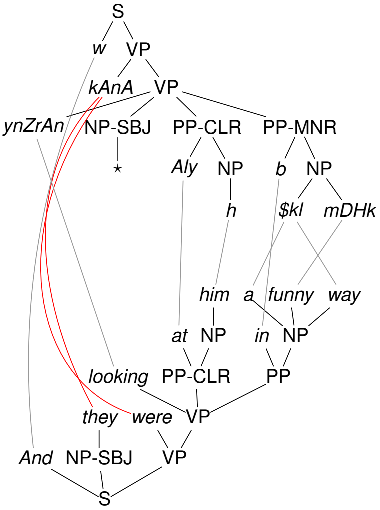
- STSSG = synch. tree-sequence substitution grammar
  [ZHANG et al., 2008] TSSG-TSSG
- ℓMBOT = local shallow multi bottom-up tree transducer
  [BRAUNE et al., 2013] LTG-TSSG

ln-XMBOT corresponds rougly to RTG-TSSG

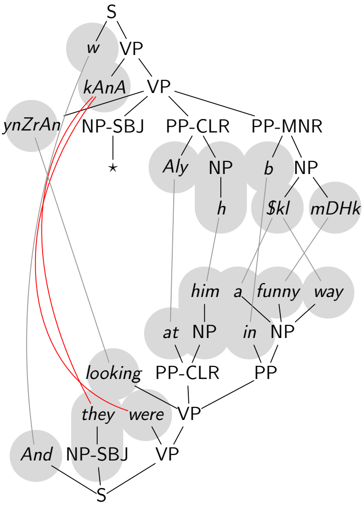
Tree Transducers in Machine Translation

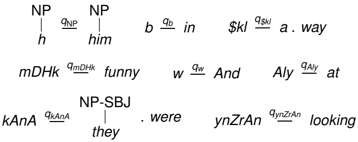
Extended Multi Bottom-up Tree Transducers

Extended Multi Bottom-up Tree Transducers

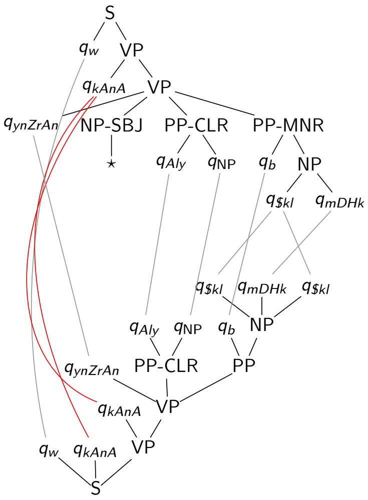


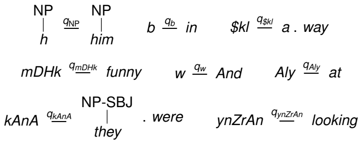
Extended Multi Bottom-up Tree Transducers



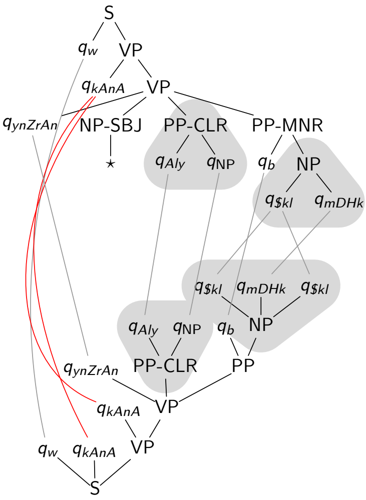


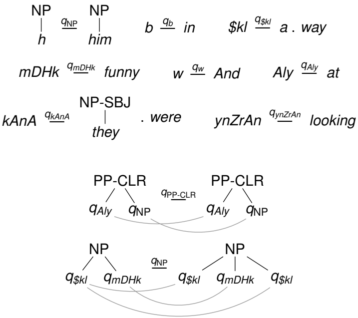
Extended Multi Bottom-up Tree Transducers



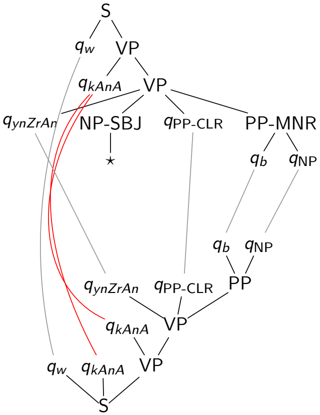


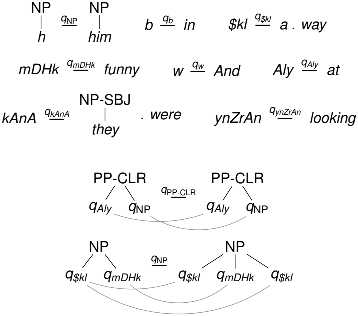
Extended Multi Bottom-up Tree Transducers



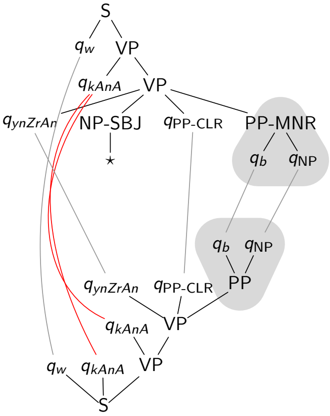


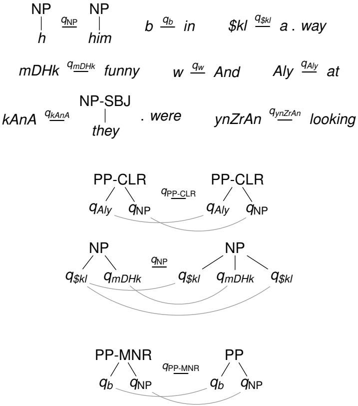
Extended Multi Bottom-up Tree Transducers





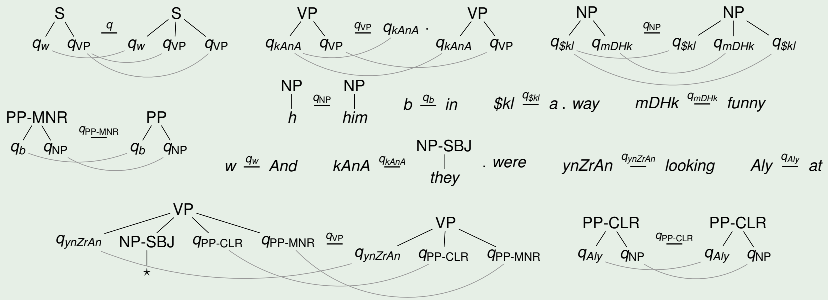
Extended Multi Bottom-up Tree Transducers





Extended Multi Bottom-up Tree Transducers

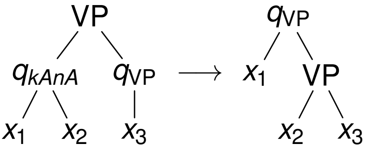
Extracted rules



Extended Multi Bottom-up Tree Transducers

Synchronous grammar notation:

Tree transducer notation (links expressed by variables):

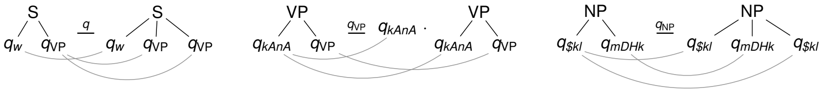


Extended Multi Bottom-up Tree Transducers

Definition (Alternative for linear XMBOT)

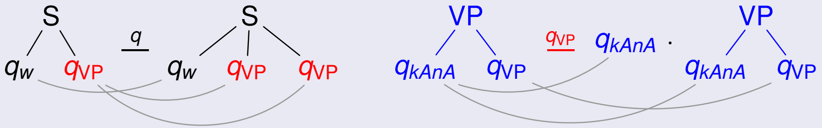
Linear extended multi bottom-up tree transducer $(Q, \Sigma, \Delta, I, R)$

- finite set Q — states
- alphabets Σ and Δ — input/output symbols
- $I \subseteq Q$ — initial states
- finite set $R \subseteq T_\Sigma(Q) \times Q \times T_\Delta(Q)^*$ — rules
  - each $q \in Q$ occurs at most once in ℓ — $(\ell, q, \vec{r}) \in R$
  - each $q \in Q$ that occurs in $\vec{r}$ also occurs in ℓ — $(\ell, q, \vec{r}) \in R$



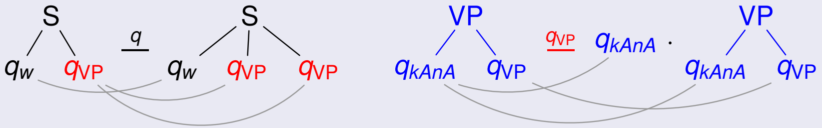
Extended Multi Bottom-up Tree Transducers

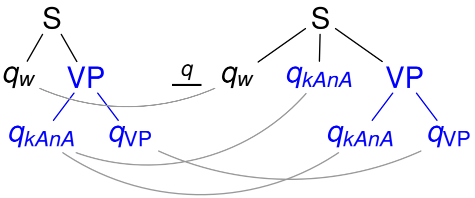
Rules



Extended Multi Bottom-up Tree Transducers

Rules





Evaluation of XMBOT

Statistical Machine Translation

Implementation [Braune et al., 2013]

- shallow ℓMBOT implemented in Moses framework
  [Koehn et al., 2007]
- variant of the syntax-based component
  [Hoang et al., 2009]

Statistical Machine Translation

Implementation [Braune et al., 2013]

- shallow ℓMBOT implemented in Moses framework [Koehn et al., 2007]
- variant of the syntax-based component [Hoang et al., 2009]
- hard- and soft-matching available
- incl. optimizations like cube pruning

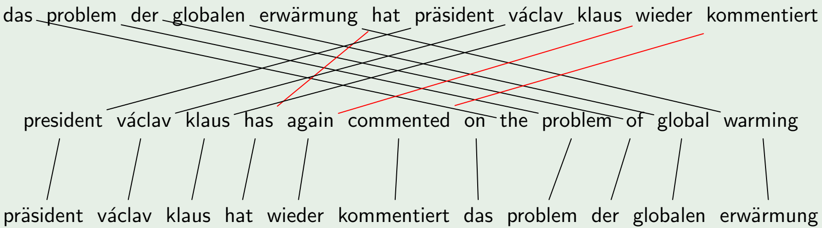
Evaluation

- English-to-German WMT 2009 translation task
- 4th version EUROPARL and news commentary (approx. 1.5 million sentence pairs)

System	BLEU-4	constrained?
University of Edinburgh (winner)	15.2	✗
GOOGLE	14.7	✗
University of Stuttgart	12.5	✗
Moses (SCFG) tree-to-tree	12.6	✓
ℓMBOT tree-to-tree	13.1	✓

Example

ℓMBOT translation



baseline translation

Statistical Machine Translation

Evaluation [Popova, 2014]

- Russian-to-English WMT'13 translation task
- Yandex corpus
  (approx. 1 million sentence pairs)

System	BLEU-4
winner [Pino et al., 2013] (hierarchical phrase-based)	25.9
hierarchical phrase-based	21.9
ℓMBOT string-to-tree	20.7
string-to-tree	19.8

Statistical Machine Translation

Pipeline

Input ⟶ Parser ⟶ Translation model ⟶ Language model ⟶ Output

Definition (One-symbol normal form)

XMBOT $(Q, \Sigma, \Delta, I, R)$ in one-symbol normal form
if ℓ contains at most one (occurrence of a) symbol of Σ

for all $(\ell, q, \vec{r}) \in R$

Definition (One-symbol normal form)

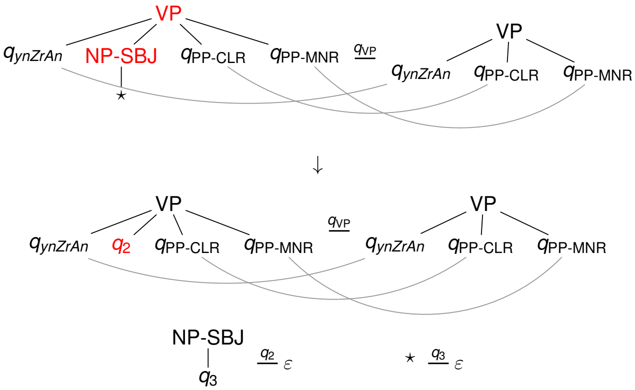
XMBOT $(Q, \Sigma, \Delta, I, R)$ in one-symbol normal form
if ℓ contains at most one (occurrence of a) symbol of Σ

for all $(\ell, q, \vec{r}) \in R$

Theorem [Engelfriet et al., 2009]

For every XMBOT there exists an equivalent XMBOT
in one-symbol normal form

Statistical Machine Translation



Transformation into one-symbol normal form

Linear-time procedure

Statistical Machine Translation

Implementation

- is often a weighted tree automaton
- yields an (unambiguous) weighted tree automaton for the parses of the input sentence
- efficient representation of $L\colon T_\Sigma \to \mathbb{R}_{\geq 0}$

Statistical Machine Translation

Implementation

- is often a weighted tree automaton
- yields an (unambiguous) weighted tree automaton for the parses of the input sentence
- efficient representation of $L \colon T_\Sigma \to \mathbb{R}_{\geq 0}$

Approach

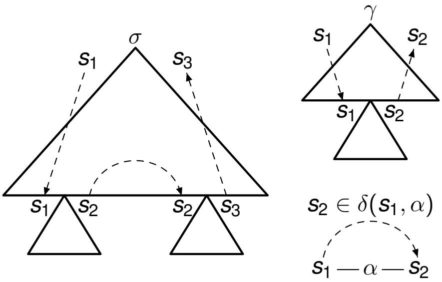
- we can intersect all (weighted) parses with our translation model (XMBOT)
- improved stability under parse errors (but only at decode)

Statistical Machine Translation

Definition (Input product)

1. weighted translation $\tau \colon T_\Sigma \times T_\Delta \to \mathbb{R}_{\geq 0}$
2. weighted language $p \colon \Sigma^* \to \mathbb{R}_{\geq 0}$ (language model)

$$_p\tau \colon T_\Sigma \times T_\Delta \to \mathbb{R}_{\geq 0} \qquad (t, u) \mapsto \tau(t, u) \cdot p(\mathrm{yd}(t))$$



Statistical Machine Translation

Definition (Input product)

1. weighted translation $\tau \colon T_\Sigma \times T_\Delta \to \mathbb{R}_{\geq 0}$
2. weighted tree language $L \colon T_\Sigma \to \mathbb{R}_{\geq 0}$ (parses)

$$_L\tau \colon T_\Sigma \times T_\Delta \to \mathbb{R}_{\geq 0} \qquad (t, u) \mapsto \tau(t, u) \cdot L(t)$$

Statistical Machine Translation

Definition (Input product)

1. weighted translation $\tau \colon T_\Sigma \times T_\Delta \to \mathbb{R}_{\geq 0}$
2. weighted tree language $L \colon T_\Sigma \to \mathbb{R}_{\geq 0}$ (parses)

$${}_L\tau \colon T_\Sigma \times T_\Delta \to \mathbb{R}_{\geq 0} \qquad (t, u) \mapsto \tau(t, u) \cdot L(t)$$

Theorem [∼, 2011]

... product of wXMBOT M with ... is

side	**wA A**	**wTA A**
input	$\mathcal{O}(\lvert M\rvert \cdot \lvert A\rvert^3)$	$\mathcal{O}(\lvert M\rvert \cdot \lvert A\rvert)$
output	$\mathcal{O}(\lvert M\rvert \cdot \lvert A\rvert^{2\,\mathrm{rk}(M)+2})$	$\mathcal{O}(\lvert M\rvert \cdot \lvert A\rvert^{\mathrm{rk}(M)})$

Statistical Machine Translation

Definition (Input product)

1. weighted translation $\tau\colon T_\Sigma \times T_\Delta \to \mathbb{R}_{\geq 0}$
2. weighted tree language $L\colon T_\Sigma \to \mathbb{R}_{\geq 0}$ (parses)

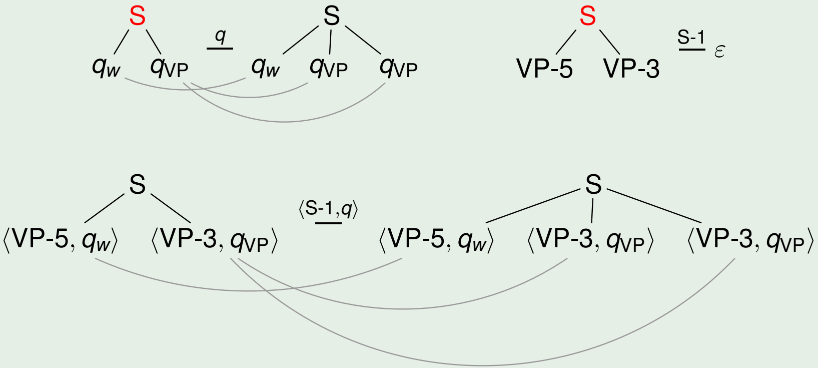
$$_L\tau\colon T_\Sigma \times T_\Delta \to \mathbb{R}_{\geq 0} \qquad (t, u) \mapsto \tau(t, u) \cdot L(t)$$

Theorem [∼, 2011]

... product of wXMBOT M with ... is

side	wA A	wTA A
input	$\mathcal{O}(\lvert M\rvert \cdot \lvert A\rvert^3)$	$\mathcal{O}(\lvert M\rvert \cdot \lvert A\rvert)$
output	$\mathcal{O}(\lvert M\rvert \cdot \lvert A\rvert^{2\,\mathrm{rk}(M)+2})$	$\mathcal{O}(\lvert M\rvert \cdot \lvert A\rvert^{\mathrm{rk}(M)})$

Example (Input product)



Statistical Machine Translation

Pipeline

Input $\longrightarrow$ Parser $\longrightarrow$ Translation model $\longrightarrow$ Language model $\longrightarrow$ Output

- exact decoding of the red part
- integrating the language model would require

$$\mathcal{O}(|M| \cdot |A|^{2\,\mathrm{rk}(M)+2})$$

Implementation [QUERNHEIM, 2014]

1. exactly computes a wTA representing the derivations for the first two models

Implementation [Quernheim, 2014]

1. exactly computes a wTA representing the derivations for the first two models
2. extracts the *k*-best derivations

Implementation [QUERNHEIM, 2014]

1. exactly computes a wTA representing the derivations for the first two models
2. extracts the *k*-best derivations
3. reranks them by the language model (i.e., multiplies their score with the LM score and resorts)

Statistical Machine Translation

Implementation [Quernheim, 2014]

1. exactly computes a wTA representing the derivations for the first two models
2. extracts the *k*-best derivations
3. reranks them by the language model (i.e., multiplies their score with the LM score and resorts)

Disadvantages

Statistical Machine Translation

Implementation [Quernheim, 2014]

1. exactly computes a wTA representing the derivations for the first two models
2. extracts the *k*-best derivations
3. reranks them by the language model
(i.e., multiplies their score with the LM score and resorts)

Disadvantages

1. s...

Statistical Machine Translation

Implementation [Quernheim, 2014]

1. exactly computes a wTA representing the derivations for the first two models
2. extracts the *k*-best derivations
3. reranks them by the language model
   (i.e., multiplies their score with the LM score and resorts)

Disadvantages

1. s. . . l. . .

Statistical Machine Translation

Implementation [Quernheim, 2014]

1. exactly computes a wTA representing the derivations for the first two models
2. extracts the *k*-best derivations
3. reranks them by the language model (i.e., multiplies their score with the LM score and resorts)

Disadvantages

1. s. . . l. . . o . . .

Statistical Machine Translation

Implementation [Quernheim, 2014]

1. exactly computes a wTA representing the derivations for the first two models
2. extracts the *k*-best derivations
3. reranks them by the language model
   (i.e., multiplies their score with the LM score and resorts)

Disadvantages

1. s. . . l. . . o . . . w

Statistical Machine Translation

Implementation [Quernheim, 2014]

1. exactly computes a wTA representing the derivations for the first two models
2. extracts the *k*-best derivations
3. reranks them by the language model
   (i.e., multiplies their score with the LM score and resorts)

Disadvantages

1. s. . . l. . . o . . . w
2. language model not integrated (needs strict structure)

Statistical Machine Translation

Implementation [QUERNHEIM, 2014]

1. exactly computes a wTA representing the derivations for the first two models
2. extracts the *k*-best derivations
3. reranks them by the language model (i.e., multiplies their score with the LM score and resorts)

Disadvantages

1. s. . . l. . . o . . . w
2. language model not integrated (needs strict structure)
3. strictness → coverage problems

Statistical Machine Translation

Evaluation

- English-to-German
- 7th EUROPARL, news commentary, and Common crawl
  (approx. 1.8–4 million sentence pairs)

System	BLEU-4	
	WMT'13	WMT'14
winner	20.8	21.0
Moses (SCFG) tree-to-tree	13.1	—
ExactMBOT tree-to-tree	16.2	17.0
Moses (SCFG) string-to-tree	14.7	—
ℓMBOT string-to-tree	15.5	—
Moses phrase-based	17.5	—

Composition

Composition

Composition

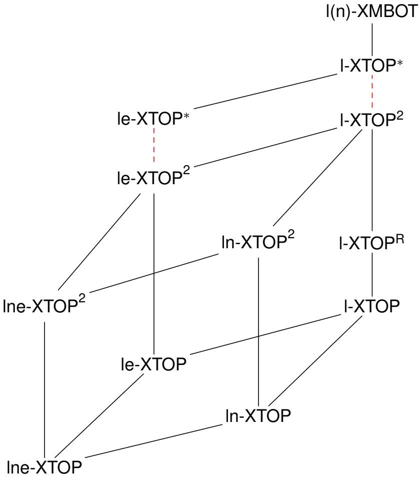
- $\tau_1 \,;\, \tau_2 = \{(s, u) \mid \exists t \colon (s, t) \in \tau_1, (t, u) \in \tau_2\}$
- support modular development
- allow integration of external knowledge sources

Question

given a class $\mathcal{C}$ of transformations, is there $n \in \mathbb{N}$ such that

$$\mathcal{C}^n = \bigcup_{k \geq 1} \mathcal{C}^k \qquad\qquad \mathcal{C}^k = \underbrace{\mathcal{C} \,;\, \cdots \,;\, \mathcal{C}}_{k \text{ times}}$$

Composition



Composition

	l-TOP	l-XTOP	l-XMBOT
ε-free, strict, nondeleting	1		1
ε-free, strict	2		1
ε-free	2		1
otherwise (without delabeling)	2		1

Composition

	l-TOP	l-XTOP	l-XMBOT
ε-free, strict, nondeleting	1	2	1
ε-free, strict	2	?	1
ε-free	2	?	1
otherwise (without delabeling)	2	?	1

Composition

e = ε-free; d = delabeling
s = strict; n = nondeleting

Theorem [Fülöp, $\sim$, 2013]

switch delabeling from back to front:

$$\text{le[s]-XTOP}^{\text{R}} \,;\, \text{l[s]d-TOP}^{\text{R}} \subseteq \text{le[s]-XTOP}^{\text{R}} \subseteq \text{l[s]d-TOP}^{\text{R}} \,;\, \text{lesn-XTOP}$$

Composition

e = ε-free; d = delabeling

s = strict; n = nondeleting

Theorem [Fülöp, ∼, 2013]

switch delabeling from back to front:

$$\text{le[s]-XTOP}^{\text{R}} \, ; \, \text{l[s]d-TOP}^{\text{R}} \subseteq \text{le[s]-XTOP}^{\text{R}} \subseteq \text{l[s]d-TOP}^{\text{R}} \, ; \, \text{lesn-XTOP}$$

Composition

e = ε-free; d = delabeling
s = strict; n = nondeleting

Theorem [Fülöp, ∼, 2013]

switch delabeling from back to front:

$$\text{le[s]-XTOP}^R \,;\, \text{l[s]d-TOP}^R \subseteq \text{le[s]-XTOP}^R \subseteq \text{l[s]d-TOP}^R \,;\, \text{lesn-XTOP}$$

Notes

- other transducer becomes strict and nondeleting

Composition

e = ε-free; d = delabeling
s = strict; n = nondeleting

Theorem [Fülöp, $\sim$, 2013]

switch delabeling from back to front:

$$\text{le[s]-XTOP}^{\text{R}} \,;\, \text{l[s]d-TOP}^{\text{R}} \subseteq \text{le[s]-XTOP}^{\text{R}} \subseteq \text{l[s]d-TOP}^{\text{R}} \,;\, \text{lesn-XTOP}$$

Notes

- other transducer becomes strict and nondeleting
- other transducer loses look-ahead

Composition

e = ε-free; d = delabeling
s = strict; n = nondeleting

Theorem

$(\text{le[s]-XTOP}^{\text{R}})^n \subseteq \text{l[s]d-TOP}^{\text{R}}$; $\text{lesn-XTOP}^2 \subseteq (\text{le[s]-XTOP}^{\text{R}})^3$

Composition

e = ε-free; d = delabeling

s = strict; n = nondeleting

Theorem

$(\text{le[s]-XTOP}^\text{R})^n \subseteq \text{l[s]d-TOP}^\text{R}$; $\text{lesn-XTOP}^2 \subseteq (\text{le[s]-XTOP}^\text{R})^3$

Proof.

$$
\begin{aligned}
& (\text{le[s]-XTOP}^\text{R})^{n+1} \\
\subseteq & \\
\subseteq & \\
\subseteq & \qquad \square
\end{aligned}
$$

Composition

e = ε-free; d = delabeling
s = strict; n = nondeleting

Theorem

$(\text{le[s]-XTOP}^{\text{R}})^n \subseteq \text{l[s]d-TOP}^{\text{R}} ; \text{lesn-XTOP}^2 \subseteq (\text{le[s]-XTOP}^{\text{R}})^3$

Proof.

$$
\begin{aligned}
& (\text{le[s]-XTOP}^{\text{R}})^{n+1} \\
\subseteq\ & \text{le[s]-XTOP}^{\text{R}} ; \text{l[s]d-TOP}^{\text{R}} ; \text{lesn-XTOP}^2 \\
\subseteq\ & \\
\subseteq\ &
\end{aligned}
$$

□

Composition

e = ε-free; d = delabeling
s = strict; n = nondeleting

Theorem

$(\text{le[s]-XTOP}^{\text{R}})^n \subseteq \text{l[s]d-TOP}^{\text{R}} \,;\, \text{lesn-XTOP}^2 \subseteq (\text{le[s]-XTOP}^{\text{R}})^3$

Proof.

$$
\begin{aligned}
& (\text{le[s]-XTOP}^{\text{R}})^{n+1} \\
\subseteq\ & \text{le[s]-XTOP}^{\text{R}} \,;\, \text{l[s]d-TOP}^{\text{R}} \,;\, \text{lesn-XTOP}^2 \\
\subseteq\ & \text{l[s]d-TOP}^{\text{R}} \,;\, \text{lesn-XTOP}^3 \\
\subseteq\ &
\end{aligned}
$$

□

Composition

e = ε-free; d = delabeling
s = strict; n = nondeleting

Theorem

$(\text{le[s]-XTOP}^{\text{R}})^n \subseteq \text{l[s]d-TOP}^{\text{R}} \,;\, \text{lesn-XTOP}^2 \subseteq (\text{le[s]-XTOP}^{\text{R}})^3$

Proof.

$$\begin{aligned} & (\text{le[s]-XTOP}^{\text{R}})^{n+1} \\ \subseteq\; & \text{le[s]-XTOP}^{\text{R}} \,;\, \text{l[s]d-TOP}^{\text{R}} \,;\, \text{lesn-XTOP}^2 \\ \subseteq\; & \text{l[s]d-TOP}^{\text{R}} \,;\, \text{lesn-XTOP}^3 \\ \subseteq\; & \text{l[s]d-TOP}^{\text{R}} \,;\, \text{lesn-XTOP}^2 \end{aligned}$$

□

Composition

Corollary

le[s]-XTOP^n $\subseteq$ QR ; l[s]d-TOP ; lesn-XTOP^2 $\subseteq$ le[s]-XTOP^4

Composition

Corollary

le[s]-XTOP^n $\subseteq$ QR ; l[s]d-TOP ; lesn-XTOP^2 $\subseteq$ le[s]-XTOP^4

Proof.

uses only standard encoding of look-ahead □

Composition

	l-TOP	le-XTOP
strict, nondeleting	1	2
strict, look-ahead	1	
strict	2	
look-ahead	1	
—	2	

Composition

	l-TOP	le-XTOP
strict, nondeleting	1	2
strict, look-ahead	1	
strict	2	
look-ahead	1	≤ 3
—	2	≤ 4

Composition

Theorem

delabeling homomorphism moving from front to back:

$$\text{lsd-HOM} \,;\, \text{les-XTOP} \subseteq \text{les-XTOP} \subseteq \text{lesn-XTOP} \,;\, \text{lsd-HOM}$$

Composition

Theorem

delabeling homomorphism moving from front to back:

$$\text{lsd-HOM} \,;\, \text{les-XTOP} \subseteq \text{les-XTOP} \subseteq \text{lesn-XTOP} \,;\, \text{lsd-HOM}$$

Notes

Composition

Theorem

delabeling homomorphism moving from front to back:

$$\text{lsd-HOM} \,;\, \text{les-XTOP} \subseteq \text{les-XTOP} \subseteq \text{les}\textcolor{red}{\text{n}}\text{-XTOP} \,;\, \text{lsd-HOM}$$

Notes

- other transducer becomes nondeleting

Composition

Theorem

delabeling homomorphism moving from front to back:

$$\text{lsd-HOM} \,;\, \text{les-XTOP} \subseteq \text{les-XTOP} \subseteq \text{lesn-XTOP} \,;\, \text{lsd-HOM}$$

Notes

- other transducer becomes nondeleting
- other transducer needs to be strict and have no look-ahead

Composition

Theorem

$(\text{les-XTOP}^{\text{R}})^n \subseteq \text{lesn-XTOP} \, ; \, \text{les-XTOP} \subseteq \text{les-XTOP}^2$

Composition

Theorem

$(\text{les-XTOP}^{\text{R}})^n \subseteq \text{lesn-XTOP} \,;\, \text{les-XTOP} \subseteq \text{les-XTOP}^2$

Proof.

$$
\begin{aligned}
(\text{les-XTOP}^{\text{R}})^{n+1} &\subseteq (\text{les-XTOP}^{\text{R}})^n \,;\, \text{les-XTOP} \\
&\subseteq \\
&\subseteq \\
&\subseteq \\
&\subseteq \\
&\subseteq
\end{aligned}
$$

□

Composition

Theorem

$(\text{les-XTOP}^{\text{R}})^n \subseteq \text{lesn-XTOP} \mathbin{;} \text{les-XTOP} \subseteq \text{les-XTOP}^2$

Proof.

$$
\begin{aligned}
(\text{les-XTOP}^{\text{R}})^{n+1} &\subseteq (\text{les-XTOP}^{\text{R}})^n \mathbin{;} \text{les-XTOP} \\
&\subseteq \text{lesn-XTOP} \mathbin{;} \text{lsd-HOM} \mathbin{;} \text{les-XTOP}^2 \\
&\subseteq \\
&\subseteq \\
&\subseteq \\
&\subseteq
\end{aligned}
$$

□

Composition

Theorem

$(\text{les-XTOP}^{\text{R}})^n \subseteq \text{lesn-XTOP} \mathbin{;} \text{les-XTOP} \subseteq \text{les-XTOP}^2$

Proof.

$$
\begin{aligned}
(\text{les-XTOP}^{\text{R}})^{n+1} &\subseteq (\text{les-XTOP}^{\text{R}})^n \mathbin{;} \text{les-XTOP} \\
&\subseteq \text{lesn-XTOP} \mathbin{;} \text{lsd-HOM} \mathbin{;} \text{les-XTOP}^2 \\
&\subseteq \text{lesn-XTOP}^3 \mathbin{;} \text{lsd-HOM} \\
&\subseteq \\
&\subseteq \\
&\subseteq
\end{aligned}
$$

□

Composition

Theorem

$(\text{les-XTOP}^{\text{R}})^n \subseteq \text{lesn-XTOP} \mathbin{;} \text{les-XTOP} \subseteq \text{les-XTOP}^2$

Proof.

$$
\begin{aligned}
(\text{les-XTOP}^{\text{R}})^{n+1} &\subseteq (\text{les-XTOP}^{\text{R}})^n \mathbin{;} \text{les-XTOP} \\
&\subseteq \text{lesn-XTOP} \mathbin{;} \text{lsd-HOM} \mathbin{;} \text{les-XTOP}^2 \\
&\subseteq \text{lesn-XTOP}^3 \mathbin{;} \text{lsd-HOM} \\
&\subseteq \text{lesn-XTOP}^2 \mathbin{;} \text{lsd-HOM} \\
&\subseteq \\
&\subseteq
\end{aligned}
$$

□

Composition

Theorem

$(\text{les-XTOP}^{\text{R}})^n \subseteq \text{lesn-XTOP} \; ; \; \text{les-XTOP} \subseteq \text{les-XTOP}^2$

Proof.

$$
\begin{aligned}
(\text{les-XTOP}^{\text{R}})^{n+1} &\subseteq (\text{les-XTOP}^{\text{R}})^n \; ; \; \text{les-XTOP} \\
&\subseteq \text{lesn-XTOP} \; ; \; \text{lsd-HOM} \; ; \; \text{les-XTOP}^2 \\
&\subseteq \text{lesn-XTOP}^3 \; ; \; \text{lsd-HOM} \\
&\subseteq \text{lesn-XTOP}^2 \; ; \; \text{lsd-HOM} \\
&\subseteq \text{lesn-XTOP} \; ; \; \text{les-XTOP}^{\text{R}} \\
&\subseteq
\end{aligned}
$$

□

Composition

Theorem

$(\text{les-XTOP}^{\text{R}})^n \subseteq \text{lesn-XTOP} \mathbin{;} \text{les-XTOP} \subseteq \text{les-XTOP}^2$

Proof.

$$
\begin{aligned}
(\text{les-XTOP}^{\text{R}})^{n+1} &\subseteq (\text{les-XTOP}^{\text{R}})^n \mathbin{;} \text{les-XTOP} \\
&\subseteq \text{lesn-XTOP} \mathbin{;} \text{lsd-HOM} \mathbin{;} \text{les-XTOP}^2 \\
&\subseteq \text{lesn-XTOP}^3 \mathbin{;} \text{lsd-HOM} \\
&\subseteq \text{lesn-XTOP}^2 \mathbin{;} \text{lsd-HOM} \\
&\subseteq \text{lesn-XTOP} \mathbin{;} \text{les-XTOP}^{\text{R}} \\
&\subseteq \text{lesn-XTOP} \mathbin{;} \text{les-XTOP}
\end{aligned}
$$

□

Composition

	l-TOP	le-XTOP
strict, nondeleting	1	2
strict, look-ahead	1	
strict	2	
look-ahead	1	≤ 3
—	2	≤ 4

Composition

	l-TOP	le-XTOP
strict, nondeleting	1	2
strict, look-ahead	1	≤ 2
strict	2	≤ 2
look-ahead	1	≤ 3
—	2	≤ 4

Composition

Definition (Hierarchy properties)

A dependency $\langle t, D, u \rangle$ is

- input hierarchical if
  1. $w_2 \not< w_1$
  2. $\exists (v_1, w_1') \in D$ with $w_1' \leq w_2$

  for all $(v_1, w_1), (v_2, w_2) \in D$ with $v_1 < v_2$

Composition

Definition (Hierarchy properties)

A dependency $\langle t, D, u \rangle$ is

- input hierarchical if
  1. $w_2 \not< w_1$
  2. $\exists(v_1, w_1') \in D$ with $w_1' \leq w_2$

  for all $(v_1, w_1), (v_2, w_2) \in D$ with $v_1 < v_2$

Composition

Definition (Hierarchy properties)

A dependency $\langle t, D, u \rangle$ is

- input hierarchical if
  1. $w_2 \not< w_1$
  2. $\exists (v_1, w_1') \in D$ with $w_1' \leq w_2$

  for all $(v_1, w_1), (v_2, w_2) \in D$ with $v_1 < v_2$
- strictly input hierarchical if
  1. $v_1 < v_2$ implies $w_1 \leq w_2$
  2. $v_1 = v_2$ implies $w_1 \leq w_2$ or $w_2 \leq w_1$

  for all $(v_1, w_1), (v_2, w_2) \in D$

Composition

Definition (Distance properties)

A dependency $\langle t, D, u \rangle$ is

- input link-distance bounded by $b \in \mathbb{N}$ if
  for all $(v_1, w_1), (v_1 v', w_2) \in D$ with $|v'| > b$
  $\exists (v_1 v, w_3) \in D$ such that $v < v'$ and $1 \leq |v| \leq b$

Composition

Definition (Distance properties)

A dependency $\langle t, D, u \rangle$ is

- input link-distance bounded by $b \in \mathbb{N}$ if
  for all $(v_1, w_1), (v_1 v', w_2) \in D$ with $|v'| > b$
  $\exists (v_1 v, w_3) \in D$ such that $v < v'$ and $1 \leq |v| \leq b$

Composition

Definition (Distance properties)

A dependency $\langle t, D, u \rangle$ is

- input link-distance bounded by $b \in \mathbb{N}$ if for all $(v_1, w_1), (v_1 v', w_2) \in D$ with $|v'| > b$ $\exists (v_1 v, w_3) \in D$ such that $v < v'$ and $1 \leq |v| \leq b$
- strict input link-distance bounded by b if for all $v_1, v_1 v' \in \text{pos}(t)$ with $|v'| > b$ $\exists (v_1 v, w_3) \in D$ such that $v < v'$ and $1 \leq |v| \leq b$

Composition

	hierarchical		link-distance bounded	
Model \ Property	input	output	input	output
l n-XTOP	strictly	strictly	strictly	strictly
l-XTOP^R	strictly	strictly	✓	strictly
l-MBOT	✓	strictly	✓	strictly

Composition

Theorem [∼ et al., 2009]

les-XTOP $\subsetneq$ les-XTOP$^{\mathrm{R}}$ $\subsetneq$ les-XTOP^2 $=$ (les-XTOP$^{\mathrm{R}}$)^2

Composition

Theorem [∼ et al., 2009]

les-XTOP $\subsetneq$ les-XTOP$^{\mathrm{R}}$ $\subsetneq$ les-XTOP^2 $=$ (les-XTOP$^{\mathrm{R}}$)^2

Proof.

- look-ahead adds power at first level
- none of the basic classes is closed under composition □

Composition

	l-TOP	le-XTOP
strict, nondeleting	1	2
strict, look-ahead	1	≤ 2
strict	2	≤ 2
look-ahead	1	≤ 3
—	2	≤ 4

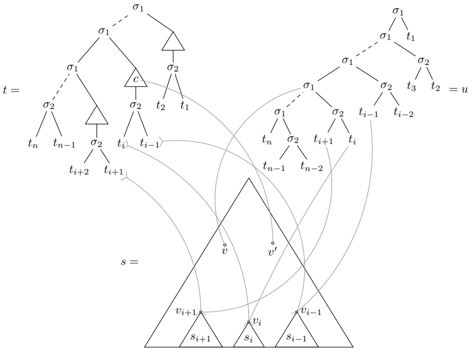
Composition

	l-TOP	le-XTOP
strict, nondeleting	1	2
strict, look-ahead	1	2
strict	2	2
look-ahead	1	≤ 3
—	2	≤ 4

Composition

Theorem [FÜLÖP, ∼, 2013]

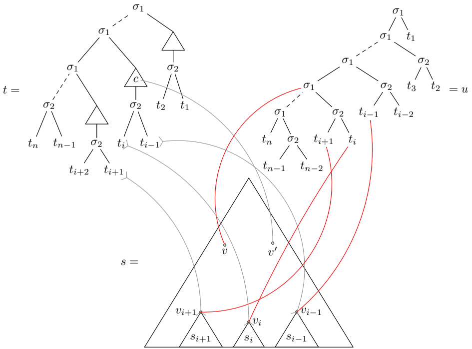
le-XTOP^2 $\subseteq$ (le-XTOP^R)^2 $\subsetneq$ le-XTOP^3 $\subseteq$ (le-XTOP^R)^3



Theorem [FÜLÖP, ∼, 2013]

le-XTOP^2 $\subseteq$ (le-XTOP^R)^2 $\subsetneq$ le-XTOP^3 $\subseteq$ (le-XTOP^R)^3

$v \not\preceq v_{i-1}$ and $v \preceq v_i$ and $v \preceq v_{i+1}$

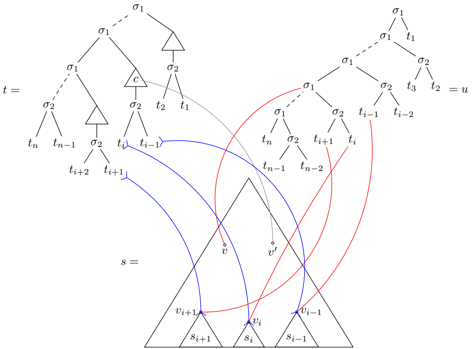


Composition

Theorem [FÜLÖP, ∼, 2013]

le-XTOP^2 $\subseteq$ (le-XTOP^R)^2 $\subsetneq$ le-XTOP^3 $\subseteq$ (le-XTOP^R)^3

$v \not\preceq v_{i-1}$ and $v \preceq v_i$ and $v \preceq v_{i+1}$

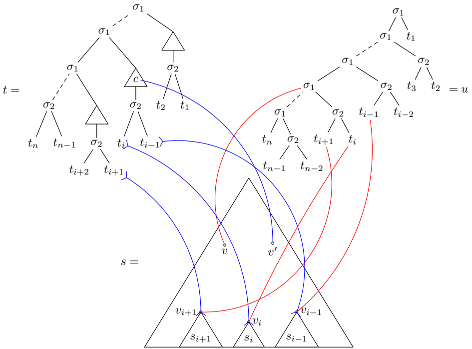


Composition

Theorem [FÜLÖP, ∼, 2013]

le-XTOP^2 $\subseteq$ (le-XTOP^R)^2 $\subsetneq$ le-XTOP^3 $\subseteq$ (le-XTOP^R)^3

$v \not\preceq v_{i-1}$ and $v \preceq v_i$ and $v \preceq v_{i+1}$
$v' \preceq v_{i-1}$ and $v' \preceq v_i$ and $v' \not\preceq v_{i+1}$



Composition

	l-TOP	le-XTOP
strict, nondeleting	1	2
strict, look-ahead	1	2
strict	2	2
look-ahead	1	≤ 3
—	2	≤ 4

Composition

	l-TOP	le-XTOP
strict, nondeleting	1	2
strict, look-ahead	1	2
strict	2	2
look-ahead	1	3
—	2	3–4 (4)

Composition

	l-TOP	l-XTOP	l-XTOP^{R}
ε-free, nondeleting	1	∞	∞
strict	2	∞	∞
nondeleting	1	∞	∞
strict, nondeleting	1	∞	∞
—	2	∞	∞

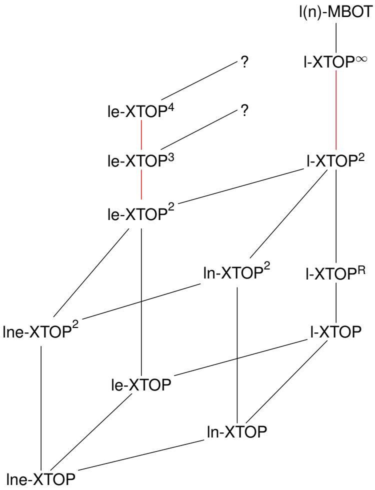
Proof.

- completely different technique [FÜLÖP, ∼, 2013] □

Composition

	l-TOP	l-XTOP	l-XTOP^R	l-MBOT
ε-free, strict, nondeleting	1	2	2	1
ε-free, strict	2	2	2	1
ε-free	2	4	3	1
otherwise (w/o delabeling)	2	∞	∞	1

Composition



Literature

Selected references

- ARNOLD, DAUCHET: *Morphismes et Bimorphismes d'Arbres* Theoret. Comput. Sci. 20, 1982
- BRAUNE, SEEMANN, QUERNHEIM, ∼: *Shallow local multi bottom-up tree transducers in statistical machine translation.* Proc. ACL, 2013
- FÜLÖP, ∼: *Composition closure of ε-free linear extended top-down tree transducers.* Proc. DLT, 2013
- ∼, GRAEHL, HOPKINS, KNIGHT: *The Power of Extended Top-down Tree Transducers.* SIAM J. Comput. 39, 2009