Synchronous Forest Substitution Grammars

Andreas Maletti

Institute for Natural Language Processing
University of Stuttgart, Germany

maletti@ims.uni-stuttgart.de

Porquerolles Island, France (CAI 2013)

Outline

Motivation

Main model

Results

Machine translation

Translation

- Input:
  Official forecasts predicted just 3 percent, Bloomberg said.

- Reference:
  Offizielle Prognosen sind von nur 3 Prozent ausgegangen, meldete Bloomberg.
  [official] [forecasts] [are] [of] [only] [3 percent] [assumed] [reported] [Bloomberg]

Machine translation

Translation

- Input:
  Official forecasts predicted just 3 percent, Bloomberg said.

- Reference:
  Offizielle Prognosen sind von nur 3 Prozent ausgegangen, meldete Bloomberg.
  [official] [forecasts] [are] [of] [only] [3 percent] [assumed] [reported] [Bloomberg]

- Google Translate (`translate.google.com`):
  Offizielle Prognosen vorhergesagt nur 3 Prozent, sagte Bloomberg.
  [official] [forecasts] [*predicted] [only] [3 percent] [said] [Bloomberg]

Machine translation

Translation

- **Input:**
  The ECB wants to hold inflation to under two percent,
  or somewhere in that vicinity.

- **Reference:**
  Die EZB ist bestrebt, die Inflationsrate unter zwei Prozent,
  [the] [ECB] [is] [desire] [the] [inflation rate] [below] [two percent]
  oder zumindest knapp an der Zwei-Prozent-Marke zu halten.
  [or] [at least] [close] [at] [the] [two percent mark] [to keep]

Machine translation

Translation

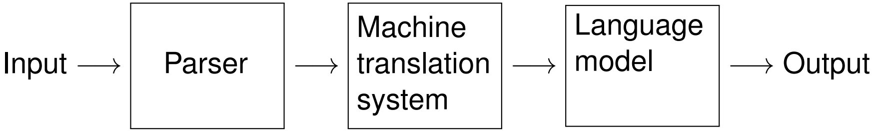
- Input:
  The ECB wants to hold inflation to under two percent,
  or somewhere in that vicinity.

- Reference:
  Die EZB ist bestrebt, die Inflationsrate unter zwei Prozent,
  [the] [ECB] [is] [desire] [the] [inflation rate] [below] [two percent]
  oder zumindest knapp an der Zwei-Prozent-Marke zu halten.
  [or] [at least] [close] [at] [the] [two percent mark] [to keep]

- Google Translate (`translate.google.com`):
  Die EZB will die Inflation unter zwei Prozent zu halten ,
  [the] [ECB] [wants] [the] [inflation] [below] [two percent] [*to keep]
  oder irgendwo in der Nähe.
  [or] [somewhere] [in] [the] [vicinity]

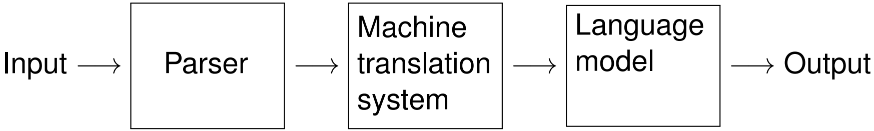
Syntax-based machine translation

Architecture



Syntax-based machine translation

Architecture



Formalisms

- Parser = weighted tree automaton
- Translation system = *some tree transducer*
- Language model = weighted string automaton

Resources

Input

- Parallel text (English and German) EUROPARL
- Parsers BITPAR, CHARNIAK, BERKELEY

Resources

Input

- Parallel text (English and German) EUROPARL
- Parsers BITPAR, CHARNIAK, BERKELEY

Example

- "We must bear in mind the Community as a whole."
- "Wir müssen uns davor hüten, alles vergemeinschaften zu wollen."

Resources

Input

- Parallel text (English and German) EUROPARL
- Parsers BITPAR, CHARNIAK, BERKELEY

Example

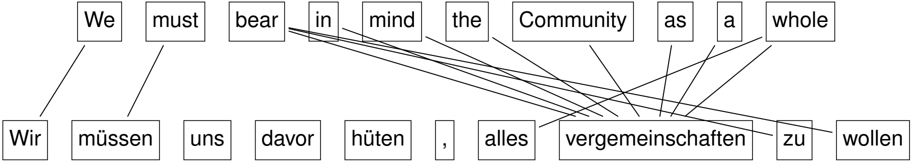
- "We must bear in mind the Community as a whole."
- "Wir müssen uns davor hüten, alles vergemeinschaften zu wollen."

EUROPARL German-English parallel data:

- 1,920,209 parallel sentences
- 44,548,491 words in German
- 47,818,827 words in English

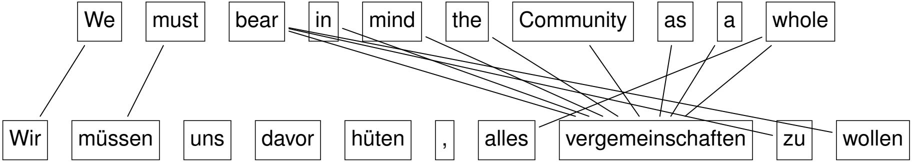
First step: word alignment

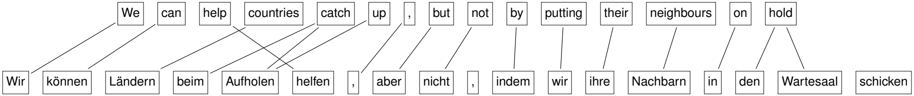
Alignments by GIZA++ [OCH, NEY '03]:



First step: word alignment

Alignments by GIZA++ [OCH, NEY '03]:

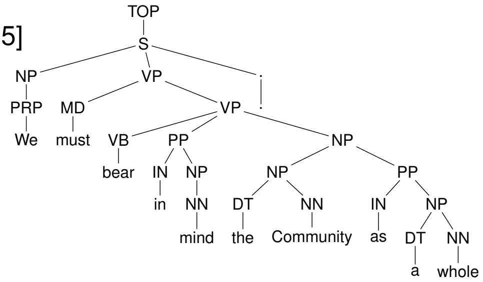




Second step: parsing

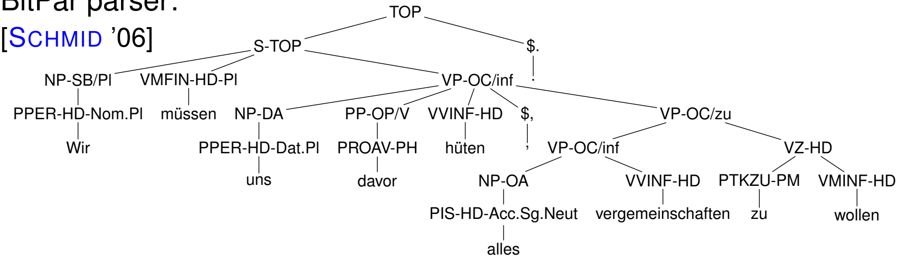
CHARNIAK parser:

[CHARNIAK, JOHNSON ’05]



BitPar parser:

[SCHMID ’06]



Full example

Parallel text

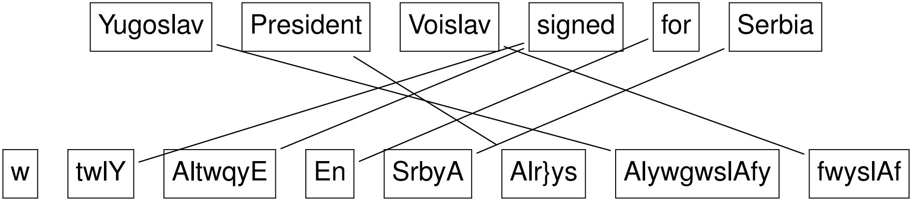
Yugoslav President Voislav signed for Serbia.
و تولى التوقيع عن صربيا الرئيس اليوغوسلافي فويسلاف
Transliteration: *w twlY AltwqyE En SrbyA Alr}ys AlywgwslAfy fwyslAf.*

And then the matter was decided, and everything was put in place.
ف كان ان تم الحسم و وضعت الأمور في نصاب ها
Transliteration: *f kAn An tm AlHsm w wDEt Al>mwr fy nSAb hA.*

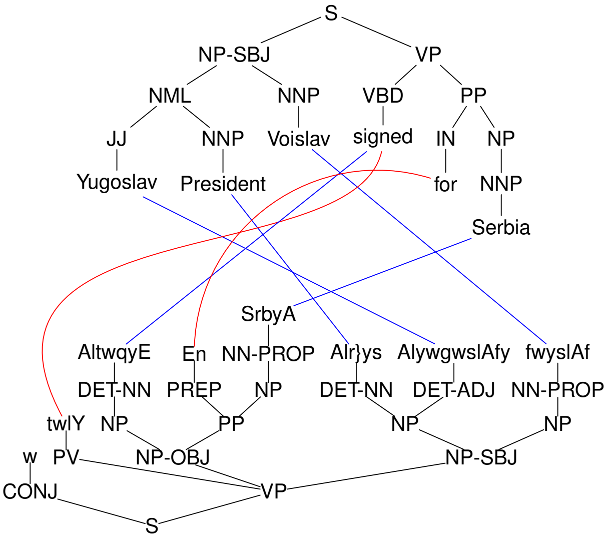
Below are the male and female winners in the different categories.
و هنا الأوائل و الأوليات في مختلف الفئات
Transliteration: *w hnA Al>wA}l w Al>wlyAt fy mxtlf Alf}At.*

Full example

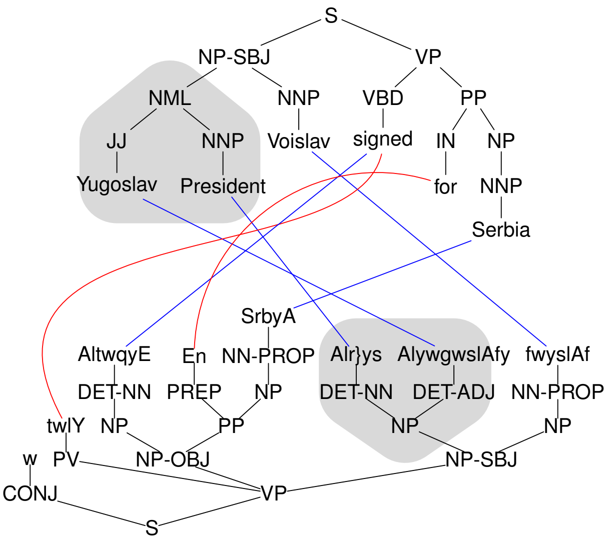
Alignment



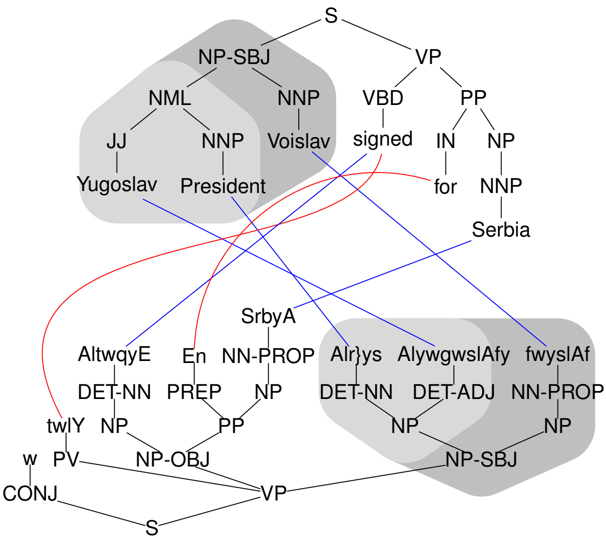
Third step: rule extraction



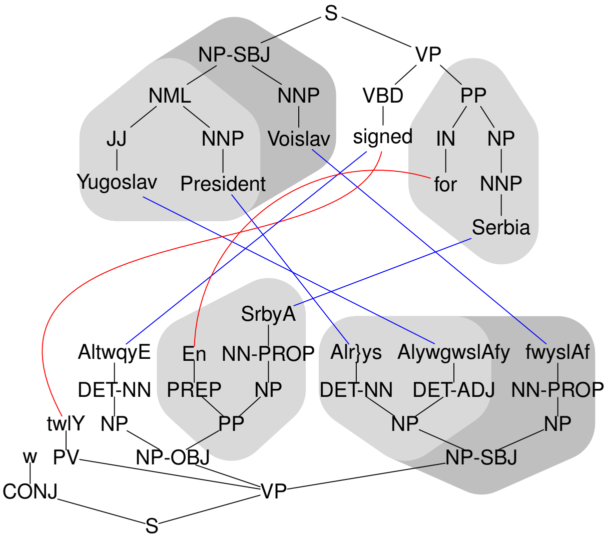
Third step: rule extraction



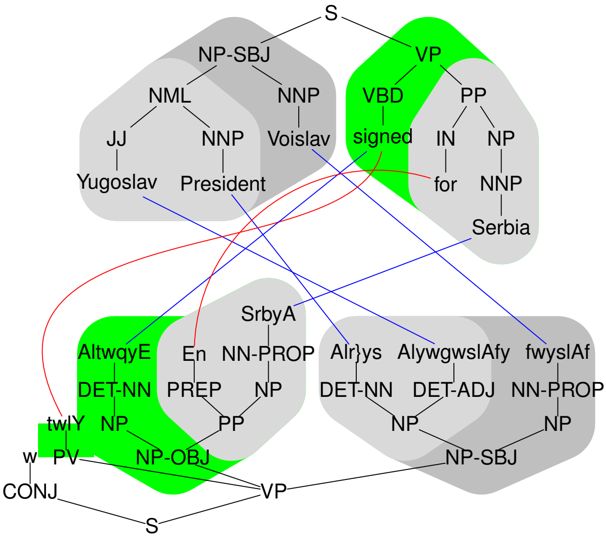
Third step: rule extraction



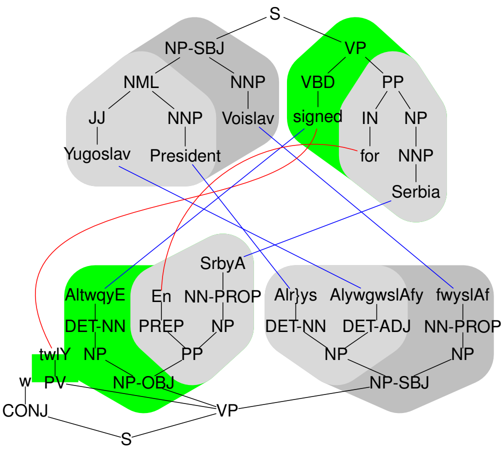
Third step: rule extraction

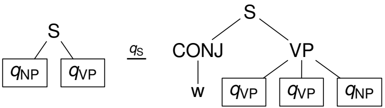


Third step: rule extraction

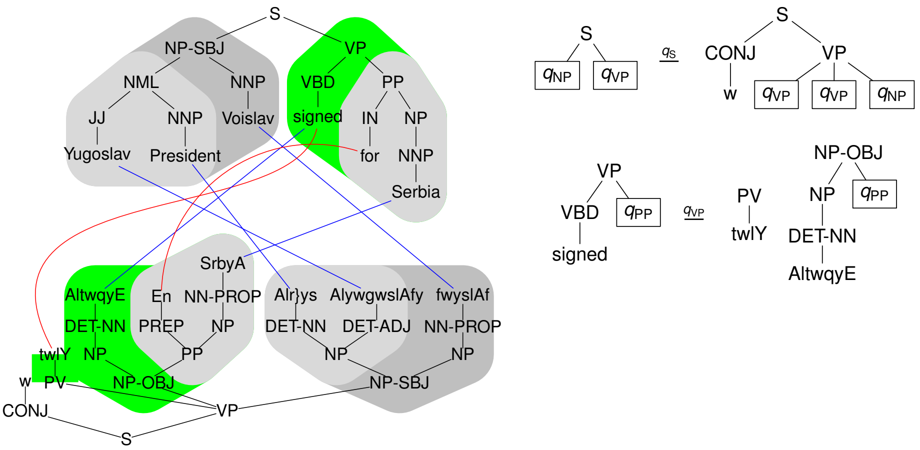


Third step: rule extraction





Third step: rule extraction



Last step: evaluation

English to German [Braune et al., 2013]

Model	BLEU
STSG (tree-to-tree)	12.60
GHKM (tree-to-string)	12.72
MBOT (tree-to-tree)	13.06

Last step: evaluation

English to German [BRAUNE et al., 2013]

Model	BLEU
STSG (tree-to-tree)	12.60
GHKM (tree-to-string)	12.72
MBOT (tree-to-tree)	13.06

Chinese to English [SUN et al., 2009]

Model	BLEU
FST (string-to-string)	23.86
STSG (tree-to-tree)	25.92
MBOT (tree-to-tree)	26.56
SFSG (tree-to-tree)	26.53

Outline

Motivation

Main model

Results

RTG — Syntax

Definition (BRAINERD, 1969)

Regular tree grammar (RTG) is tuple $G = (Q, \Sigma, I, P)$

- alphabet Q — *nonterminals*
- alphabet Σ — *terminals*
- $I \subseteq Q$ — *initial nonterminals*
- finite set $P \subseteq Q \times T_\Sigma(Q)$ — *productions*

RTG — Syntax

Definition (BRAINERD, 1969)

Regular tree grammar (RTG) is tuple $G = (Q, \Sigma, I, P)$

- alphabet Q *nonterminals*
- alphabet Σ *terminals*
- $I \subseteq Q$ *initial nonterminals*
- finite set $P \subseteq Q \times T_\Sigma(Q)$ *productions*

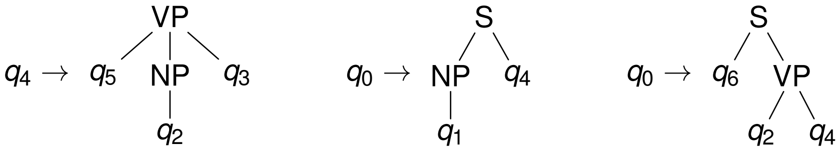
Remark

Instead of (q, r) we write $q \to r$

RTG — Syntax

Example

- $Q = \{q_0, q_1, q_2, q_3, q_4, q_5, q_6\}$
- $\Sigma = \{\mathrm{VP}, \mathrm{NP}, \mathrm{S}, \dots\}$
- $I = \{q_0\}$
- and the following productions:



RTG — Semantics

Definition (Derivation semantics)

Sentential forms: $t, u \in T_\Sigma(Q)$

$$t \Rightarrow_G u$$

if there exist position $w \in \text{pos}(t)$ and production $q \to r \in P$

- $t = t[q]_w$
- $u = t[r]_w$

RTG — Semantics

Definition (Derivation semantics)

Sentential forms: $t, u \in T_\Sigma(Q)$

$$t \Rightarrow_G u$$

if there exist position $w \in \text{pos}(t)$ and production $q \to r \in P$

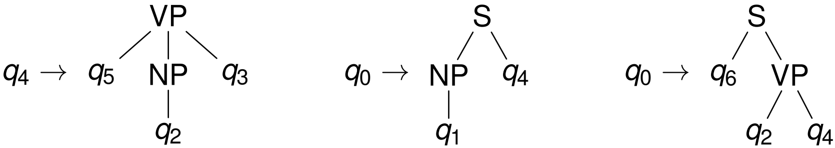
- $t = t[q]_w$
- $u = t[r]_w$

Definition (Recognized tree language)

$$L(G) = \{t \in T_\Sigma \mid \exists q \in I \colon q \Rightarrow_G^* t\}$$

RTG — Semantics

Example (Productions)

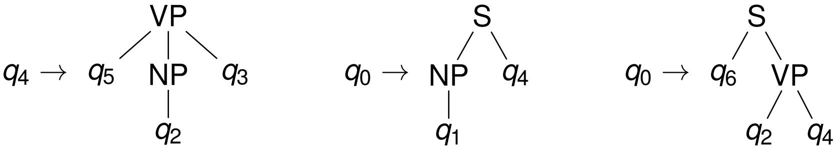


Example (Derivation)

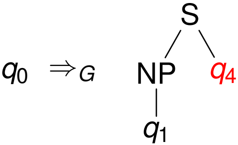
q_0

RTG — Semantics

Example (Productions)

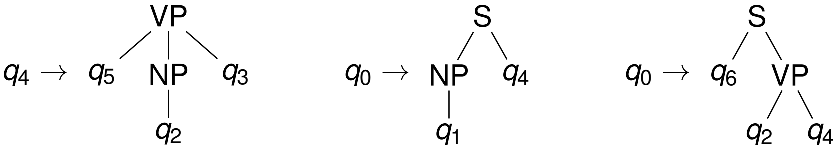


Example (Derivation)

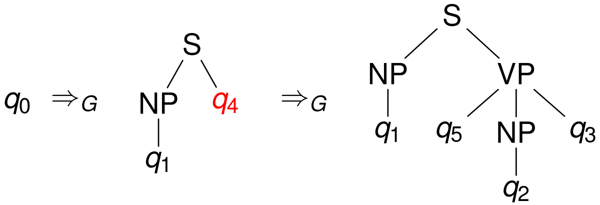


RTG — Semantics

Example (Productions)



Example (Derivation)



Synchronous Grammar

Intuition

- Productions that consist of several RTG productions
- Synchronous = several RTG productions are applied at once
- typically at least two RTG productions *input/output side*

Synchronous Grammar

Intuition

- Productions that consist of several RTG productions
- Synchronous = several RTG productions are applied at once
- typically at least two RTG productions *input/output side*
- but even several RTG productions per side possible

Synchronous Grammar

Intuition

- Productions that consist of several RTG productions
- Synchronous = several RTG productions are applied at once
- typically at least two RTG productions *input/output side*
- but even several RTG productions per side possible

Definition

Given RTG (Q, Σ, I, P) and $q \in Q$ let

$$P_q = \{q \to r \mid q \to r \in P\}$$

SFSG — Syntax

Definition (RAOULT, 1997 and SUN et al., 2009)

Synchronous forest substitution grammar (SFSG) is tuple

$$G = (Q, \Sigma, I, P, R)$$

- (Q, Σ, I, P) is RTG *basic productions*
- $R \subseteq (\bigcup_{q \in I} P_q \times P_q) \cup (\bigcup_{q \in Q \setminus I} P_q^* \times P_q^*)$ finite *aligned rules*

SFSG — Syntax

Definition (Raoult, 1997 and Sun et al., 2009)

Synchronous forest substitution grammar (SFSG) is tuple

$$G = (Q, \Sigma, I, P, R)$$

- (Q, Σ, I, P) is RTG *basic productions*
- $R \subseteq (\bigcup_{q \in I} P_q \times P_q) \cup (\bigcup_{q \in Q \setminus I} P_q^* \times P_q^*)$ finite *aligned rules*

Definition (Arnold & Dauchet, 1982)

Multi bottom-up tree transducer (MBOT) is SFSG with $R \subseteq \bigcup_{q \in Q} P_q \times P_q^*$

SFSG — Syntax

Example

RTG $(Q, \Sigma, \{q_0\}, P)$

- $Q = \{q_0, q, q'\}$ and $\Sigma = \{\alpha, \gamma_1, \gamma_2, \sigma\}$
- P contains the productions:

$$\begin{array}{llll} \rho_0 : q_0 \to \sigma(q, q', q) & \rho_2 : q \to \gamma_1(q) & \rho_4 : q \to \gamma_2(q) & \rho_6 : q \to \alpha \\ \rho_1 : q_0 \to \sigma(q', \alpha, q') & \rho_3 : q' \to \gamma_1(q') & \rho_5 : q' \to \gamma_2(q') & \rho_7 : q' \to \alpha \end{array}$$

SFSG — Syntax

Example

RTG $(Q, \Sigma, \{q_0\}, P)$

- $Q = \{q_0, q, q'\}$ and $\Sigma = \{\alpha, \gamma_1, \gamma_2, \sigma\}$
- P contains the productions:

$$\begin{array}{llll} \rho_0 : q_0 \to \sigma(q, q', q) & \rho_2 : q \to \gamma_1(q) & \rho_4 : q \to \gamma_2(q) & \rho_6 : q \to \alpha \\ \rho_1 : q_0 \to \sigma(q', \alpha, q') & \rho_3 : q' \to \gamma_1(q') & \rho_5 : q' \to \gamma_2(q') & \rho_7 : q' \to \alpha \end{array}$$

SFSG $G = (Q, \Sigma, \{q_0\}, P, R)$

$$R = \{(\rho_0, \rho_1), (\rho_2\rho_2, \varepsilon), (\rho_4\rho_4, \varepsilon), (\rho_6\rho_6, \varepsilon), (\rho_3, \rho_3\rho_3), (\rho_5, \rho_5\rho_5)\,(\rho_7, \rho_7\rho_7)\}$$

SFSG — Syntax

Example

RTG $(Q, \Sigma, \{q_0\}, P)$

- $Q = \{q_0, q, q'\}$ and $\Sigma = \{\alpha, \gamma_1, \gamma_2, \sigma\}$
- P contains the productions:

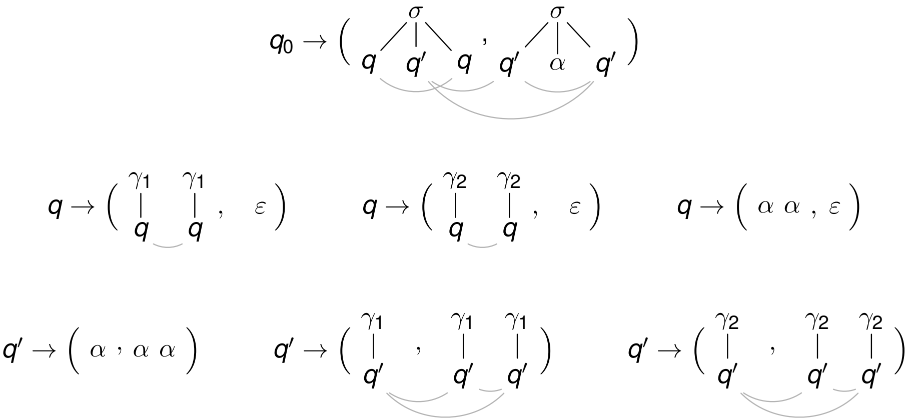
$$\begin{array}{llll} \rho_0 : q_0 \to \sigma(q, q', q) & \rho_2 : q \to \gamma_1(q) & \rho_4 : q \to \gamma_2(q) & \rho_6 : q \to \alpha \\ \rho_1 : q_0 \to \sigma(q', \alpha, q') & \rho_3 : q' \to \gamma_1(q') & \rho_5 : q' \to \gamma_2(q') & \rho_7 : q' \to \alpha \end{array}$$

SFSG $G = (Q, \Sigma, \{q_0\}, P, R)$

$$R = \{(\rho_0, \rho_1), (\rho_2\rho_2, \varepsilon), (\rho_4\rho_4, \varepsilon), (\rho_6\rho_6, \varepsilon), (\rho_3, \rho_3\rho_3), (\rho_5, \rho_5\rho_5)\, (\rho_7, \rho_7\rho_7)\}$$

G is no MBOT

SFSG — Syntax



SFSG — Semantics

Definition

Pre-translation for SFSG $G = (Q, \Sigma, I, P, R)$ is triple $\langle \vec{t}, q, \vec{u} \rangle$

- $q \in Q$ *(governing) nonterminal*
- $\vec{t}, \vec{u} \in T_\Sigma^*$ *input and output tree sequences*

Definition

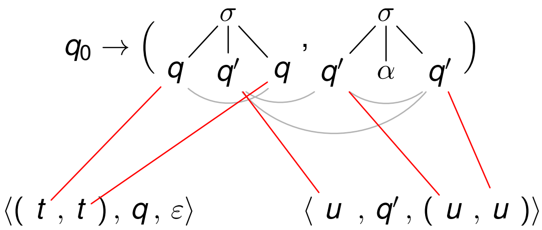
Pre-translations $\mathrm{PT}(G)$ generated by G are smallest set T

$$\langle \vec{\ell}\theta \, , \, q \, , \, \vec{r}\theta' \rangle \in \mathrm{PT}(G)$$

for all $\chi = q \to (\vec{\ell}, \vec{r}) \in R$, maps $\theta, \theta' \colon \mathrm{var}(\chi) \to T_\Sigma^*$ and $q' \in \mathrm{var}(\chi)$

- $|\theta(q')| = |\mathrm{pos}_{q'}(\vec{\ell})|$ and $|\theta'(q')| = |\mathrm{pos}_{q'}(\vec{r})|$
- $\langle \theta(q'), q', \theta'(q') \rangle \in T$

SFSG — Semantics



SFSG — Semantics

Definition

SFSG $G = (Q, \Sigma, I, P, R)$ computes the tree translation $\tau_G \subseteq T_\Sigma \times T_\Sigma$

$$\tau_G = \bigcup_{q \in I} \{(t, u) \mid \langle t, q, u \rangle \in \mathrm{PT}(G)\}$$

SFSG — Semantics

Definition

SFSG $G = (Q, \Sigma, I, P, R)$ computes the tree translation $\tau_G \subseteq T_\Sigma \times T_\Sigma$

$$\tau_G = \bigcup_{q \in I} \{(t, u) \mid \langle t, q, u \rangle \in \mathrm{PT}(G)\}$$

Definition

- SFSG = translations computable by SFSG
- MBOT = translations computable by MBOT

Outline

Motivation

Main model

Results

One-symbol normal form

Definition

MBOT (Q, Σ, I, P, R) is in one-symbol normal form
if $|\mathrm{pos}_\Sigma(\ell)| \leq 1$ for every $q \to (\ell, \vec{r}) \in R$

One-symbol normal form

Definition

MBOT (Q, Σ, I, P, R) is in one-symbol normal form
if $|\text{pos}_\Sigma(\ell)| \leq 1$ for every $q \to (\ell, \vec{r}) \in R$

Lemma (ENGELFRIET et al., 2009)

Every MBOT can be transformed into one-symbol normal form

Simple properties

Theorem

1. $\mathrm{SFSG} = \mathrm{SFSG}^{-1}$
2. *Domain* $\text{dom}(\tau)$ *and range* $\text{ran}(\tau)$ *of* $\tau \in \mathrm{SFSG}$ *are not necessarily regular*
3. $\mathrm{MBOT} \subsetneq \mathrm{SFSG}$

Decomposition

Theorem (RAOULT, 1997)

For every SFSG G there exist two (deterministic) MBOT G_1 and G_2

$$\tau_G = \tau_{G_1}^{-1} ; \tau_{G_2}$$

Decomposition

Theorem (RAOULT, 1997)

For every SFSG G there exist two (deterministic) MBOT G_1 and G_2

$$\tau_G = \tau_{G_1}^{-1} \,;\, \tau_{G_2}$$

Theorem (Bimorphism characterization)

$$\text{SFSG} = \text{d-MBOT}^{-1} \,;\, \text{RTG} \,;\, \text{d-MBOT}$$

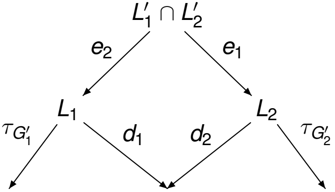
Composition

Theorem

$$\text{MBOT}^{-1} \,;\, \text{MBOT} \subseteq \text{SFSG}$$

Proof.

- Decompose MBOT into bimorphisms $(d_1, L_1, \tau_{G'_1})$ and $(d_2, L_2, \tau_{G'_2})$
- Apply



- Use SFSG bimorphism characterization □

Characterization

Corollary

$$\text{SFSG} = \text{MBOT}^{-1} \; ; \text{MBOT}$$

Immediate consequences

Theorem (Radmacher, 2008)

SFSG *is not closed under composition*

Corollary

$\mathrm{MBOT} \,;\, \mathrm{MBOT}^{-1} \not\subseteq \mathrm{SFSG}$.

Immediate consequences

Theorem (Radmacher, 2008)

SFSG *is not closed under composition*

Corollary

$\text{MBOT} \mathbin{;} \text{MBOT}^{-1} \not\subseteq \text{SFSG}$.

Proof.

$$\text{SFSG} \mathbin{;} \text{SFSG} \subseteq (\text{MBOT}^{-1} \mathbin{;} \text{MBOT}) \mathbin{;} (\text{MBOT}^{-1} \mathbin{;} \text{MBOT})$$

□

Immediate consequences

Theorem (RADMACHER, 2008)

SFSG *is not closed under composition*

Corollary

$\text{MBOT} ; \text{MBOT}^{-1} \not\subseteq \text{SFSG}$.

Proof.

$$\text{SFSG} ; \text{SFSG} \subseteq (\text{MBOT}^{-1} ; \text{MBOT}) ; (\text{MBOT}^{-1} ; \text{MBOT})$$
$$\subseteq \text{MBOT}^{-1} ; \text{SFSG} ; \text{MBOT}$$

□

Immediate consequences

Theorem (RADMACHER, 2008)

SFSG *is not closed under composition*

Corollary

$\text{MBOT} \mathbin{;} \text{MBOT}^{-1} \not\subseteq \text{SFSG}$.

Proof.

$$\begin{aligned}\text{SFSG} \mathbin{;} \text{SFSG} &\subseteq (\text{MBOT}^{-1} \mathbin{;} \text{MBOT}) \mathbin{;} (\text{MBOT}^{-1} \mathbin{;} \text{MBOT}) \\ &\subseteq \text{MBOT}^{-1} \mathbin{;} \text{SFSG} \mathbin{;} \text{MBOT} \\ &\subseteq \text{MBOT}^{-1} \mathbin{;} (\text{MBOT}^{-1} \mathbin{;} \text{MBOT}) \mathbin{;} \text{MBOT}\end{aligned}$$

□

Immediate consequences

Theorem (RADMACHER, 2008)

SFSG *is not closed under composition*

Corollary

$\mathrm{MBOT} \mathbin{;} \mathrm{MBOT}^{-1} \not\subseteq \mathrm{SFSG}$.

Proof.

$$\begin{aligned}
\mathrm{SFSG} \mathbin{;} \mathrm{SFSG} &\subseteq (\mathrm{MBOT}^{-1} \mathbin{;} \mathrm{MBOT}) \mathbin{;} (\mathrm{MBOT}^{-1} \mathbin{;} \mathrm{MBOT}) \\
&\subseteq \mathrm{MBOT}^{-1} \mathbin{;} \mathrm{SFSG} \mathbin{;} \mathrm{MBOT} \\
&\subseteq \mathrm{MBOT}^{-1} \mathbin{;} (\mathrm{MBOT}^{-1} \mathbin{;} \mathrm{MBOT}) \mathbin{;} \mathrm{MBOT} \\
&\subseteq \mathrm{MBOT}^{-1} \mathbin{;} \mathrm{MBOT}
\end{aligned}$$

□

Immediate consequences

Theorem (RADMACHER, 2008)

SFSG *is not closed under composition*

Corollary

$\mathrm{MBOT} ; \mathrm{MBOT}^{-1} \not\subseteq \mathrm{SFSG}$.

Proof.

$$\begin{aligned} \mathrm{SFSG} ; \mathrm{SFSG} &\subseteq (\mathrm{MBOT}^{-1} ; \mathrm{MBOT}) ; (\mathrm{MBOT}^{-1} ; \mathrm{MBOT}) \\ &\subseteq \mathrm{MBOT}^{-1} ; \mathrm{SFSG} ; \mathrm{MBOT} \\ &\subseteq \mathrm{MBOT}^{-1} ; (\mathrm{MBOT}^{-1} ; \mathrm{MBOT}) ; \mathrm{MBOT} \\ &\subseteq \mathrm{MBOT}^{-1} ; \mathrm{MBOT} \\ &= \mathrm{SFSG} \end{aligned}$$

□

Immediate consequences

problem	string level	tree level
Parsing	$\mathcal{O}(\lvert G\rvert \cdot (\lvert w_1\rvert \cdot \lvert w_2\rvert)^{2\,\mathrm{rk}(G)+2})$	$\mathcal{O}(\lvert G\rvert \cdot \lvert t_1\rvert \cdot \lvert t_2\rvert)$
Translation	$\mathcal{O}(\lvert G\rvert \cdot \lvert w_1\rvert^{2\,\mathrm{rk}(G)+2})$	$\mathcal{O}(\lvert G\rvert \cdot \lvert t_1\rvert)$