Compositions of Weighted Extended Top-down Tree Transducers

Andreas Maletti

Faculty of Mathematics and Computer Science
Universität Leipzig, Germany
maletti@informatik.uni-leipzig.de

Eger, Hungary — September 7, 2023

Composition

Build translation system  $\rightarrow$  (no suitable corpus)

English to Catalan

Composition

Build translation system  →  (no suitable corpus)

English to Catalan

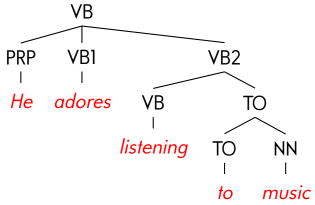
Pivot approach via Composition

 →  → 

English to Spanish to Catalan

Composition

Modular approach to translation system [Knight, Graehl, CoLing 2005]



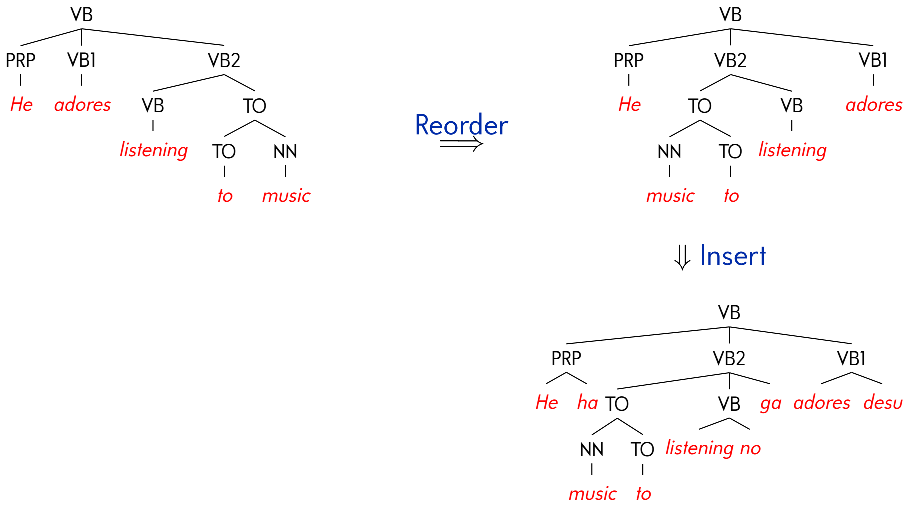
Composition

Modular approach to translation system

[Knight, Graehl, CoLing 2005]

Modular approach to translation system

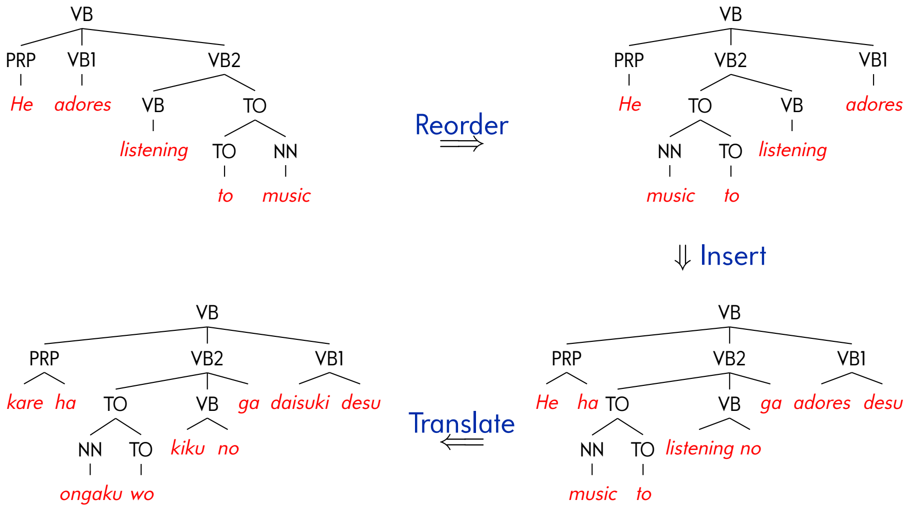
[Knight, Graehl, CoLing 2005]



Composition

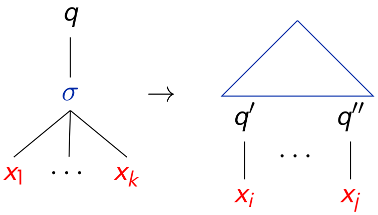
Modular approach to translation system

[Knight, Graehl, CoLing 2005]



Top-down Tree Transducer

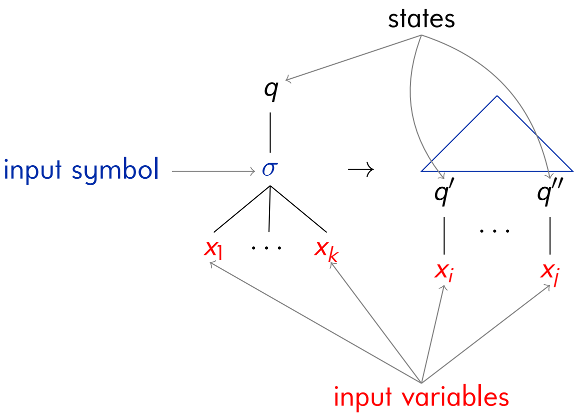
Rule shape [Rounds 1968], [Thatcher 1970]



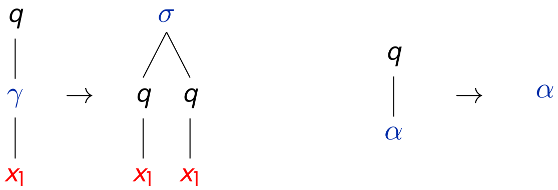
Top-down Tree Transducer

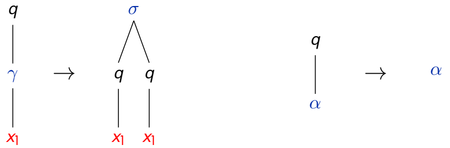
Rule shape

[Rounds 1968], [Thatcher 1970]



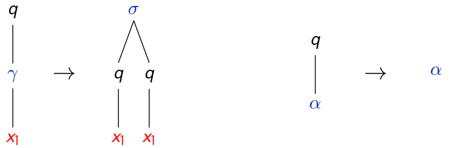
Example rules



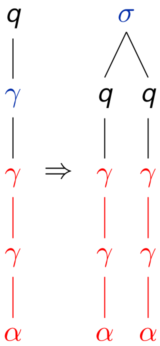


Example derivation

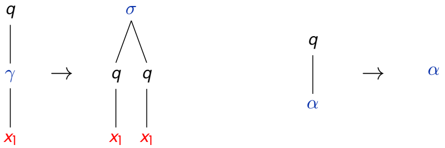
Top-down Tree Transducer



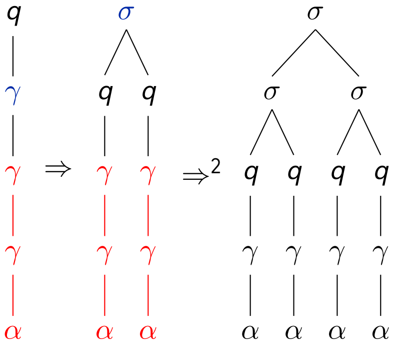
Example derivation



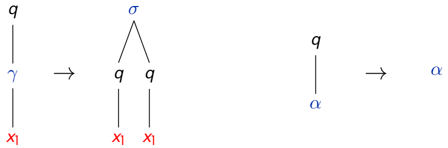
Top-down Tree Transducer



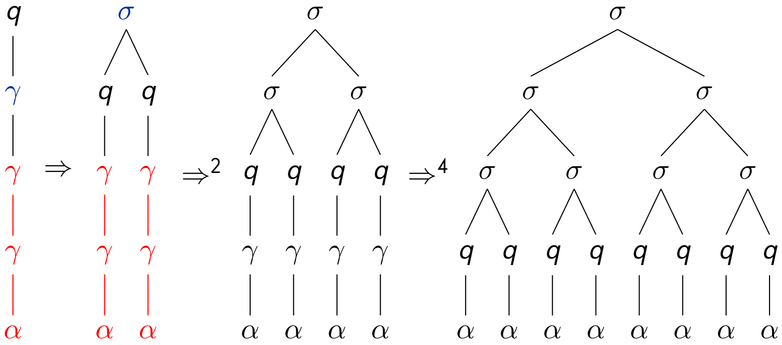
Example derivation



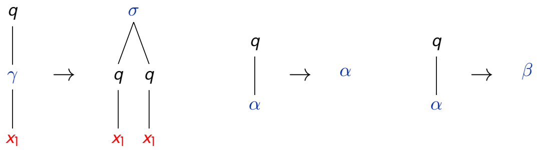
Top-down Tree Transducer



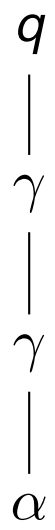
Example derivation

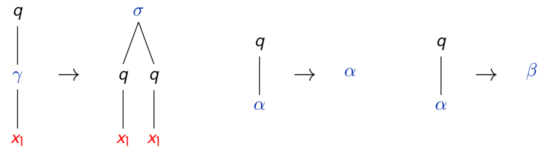


Top-down Tree Transducer

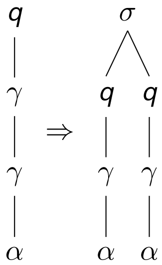


Example derivation

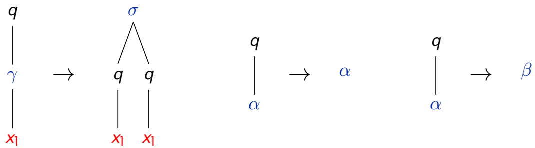




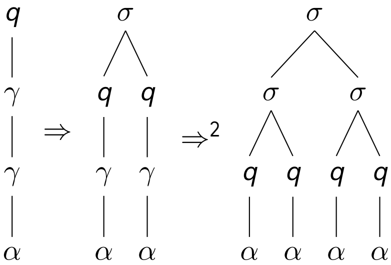
Example derivation

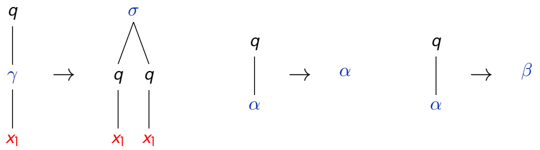


Top-down Tree Transducer

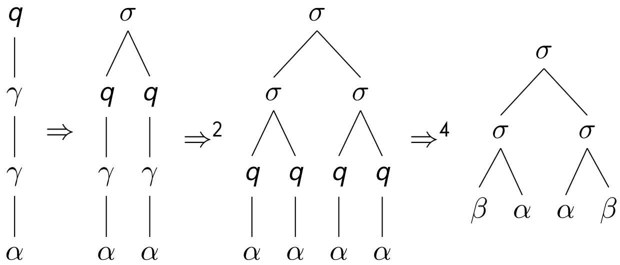


Example derivation

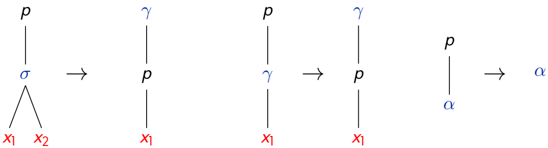




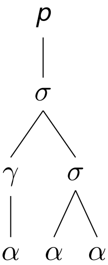
Example derivation

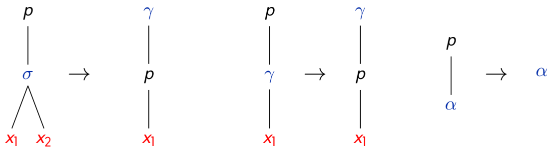


Top-down Tree Transducer

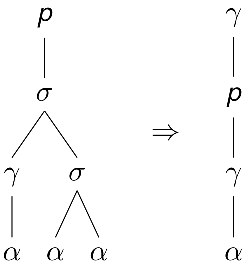


Example derivation

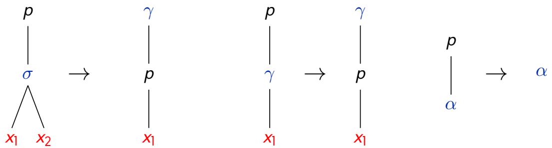




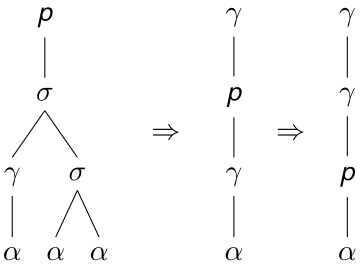
Example derivation



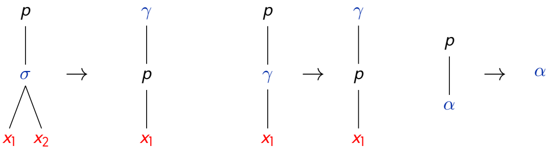
Top-down Tree Transducer



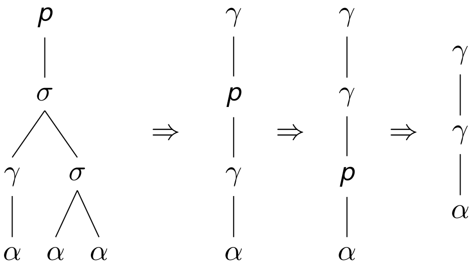
Example derivation



Top-down Tree Transducer



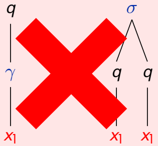
Example derivation



Top-down Tree Transducer

Definition

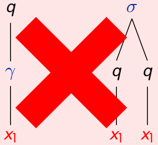
- linear if no repeat variable in rhs



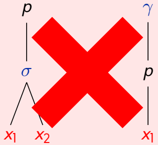
Top-down Tree Transducer

Definition

- linear if no repeat variable in rhs

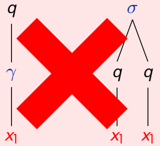


- nondeleting if every variable of lhs occurs in rhs

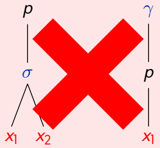


Definition

- linear if no repeat variable in rhs



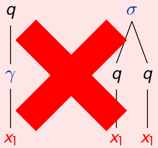
- nondeleting if every variable of lhs occurs in rhs



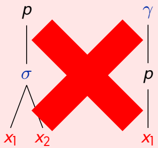
- deterministic if lhs uniquely determines rhs
  (and at most 1 initial state)

Definition

- linear if no repeat variable in rhs

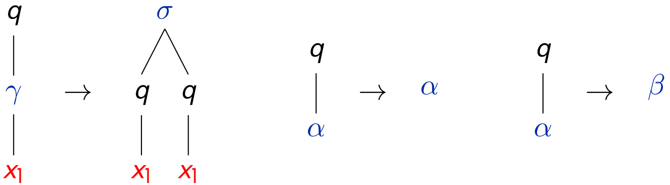


- nondeleting if every variable of lhs occurs in rhs

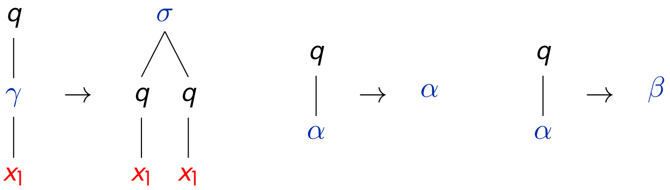


- deterministic if lhs uniquely determines rhs
  (and at most 1 initial state)
- total if at least one rule exists for every lhs
  (and at least 1 initial state)

Example

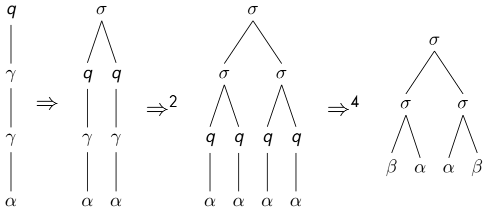


Example



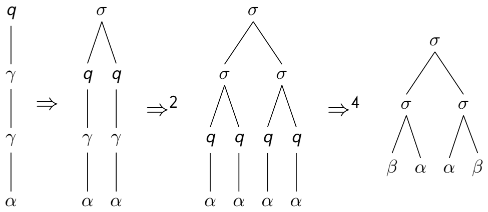
not linear (copying), nondeleting, not deterministic, and total

Translation of 1st transducer

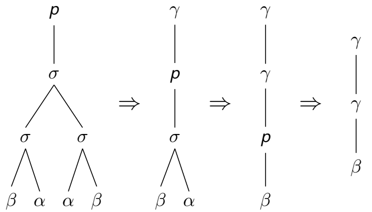


General Composition Strategy

Translation of 1st transducer



Translation of 2nd transducer

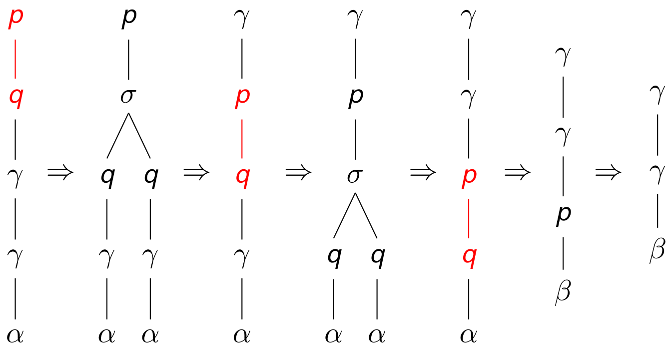


General Composition Strategy

1. Start both transducers on input (in correct order)
2. Apply derivation step of 1st transducer
3. Apply derivation steps of 2nd transducer as long as possible
   (to immediately consume intermediate trees)

General Composition Strategy

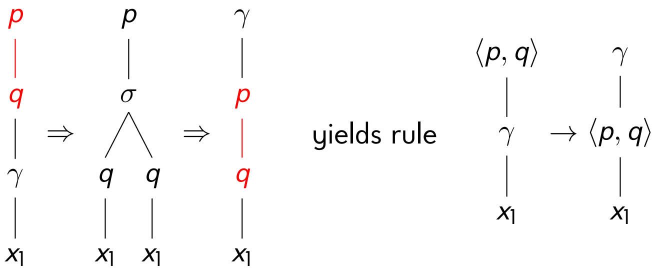
Interspersed translation



General Composition Strategy

- Strategy also works on rule level
  (combine single rule of 1st transducer + some rules of 2nd transducer)
- Pair states $\begin{matrix} p \\ | \\ q \end{matrix}$ to simply $\langle p, q \rangle$
- Yields classical composition construction
- Not universally correct, but yields celebrated composition results

- Strategy also works on rule level
  (combine single rule of 1st transducer + some rules of 2nd transducer)
- Pair states $\genfrac{}{}{0pt}{}{p}{q}$ to simply $\langle p, q\rangle$
- Yields classical composition construction
- Not universally correct, but yields celebrated composition results



General Composition Strategy

Composition results [Engelfriet 1975], [Baker 1979]

Case	1st transducer	2nd transducer
(a)		linear and nondeleting
(b)	total	linear
(c)	deterministic	nondeleting
(d)	deterministic and total	

General Composition Strategy

Composition results [Engelfriet 1975], [Baker 1979]

Case	1st transducer	2nd transducer
(a)		linear and nondeleting
(b)	total	linear
(c)	deterministic	nondeleting
(d)	deterministic and total	

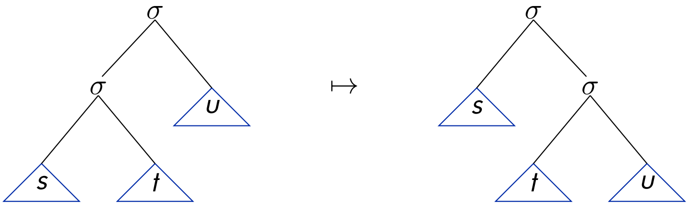
Problem

- Transducers in NLP usually not deterministic — excluding (c) and (d)
- Rotation requires nonlinear transducers — excluding (a) and (b)
- Rotation is relevant linguistic reordering

Problem

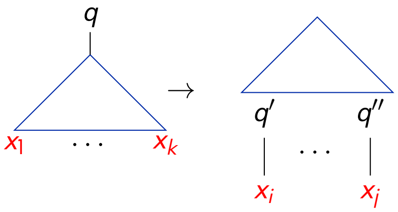
- Transducers in NLP usually not deterministic — excluding (c) and (d)
- Rotation requires nonlinear transducers — excluding (a) and (b)
- Rotation is relevant linguistic reordering

Rotation



Extended Top-down Tree Transducer

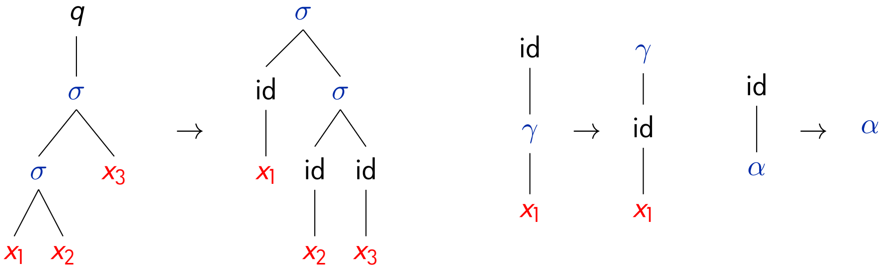
Rule shape [Arnold, Dauchet 1976], [∼, Graehl, Hopkins, Knight 2009]



(Disallow lhs $\underset{x_1}{\overset{q}{|}}$)

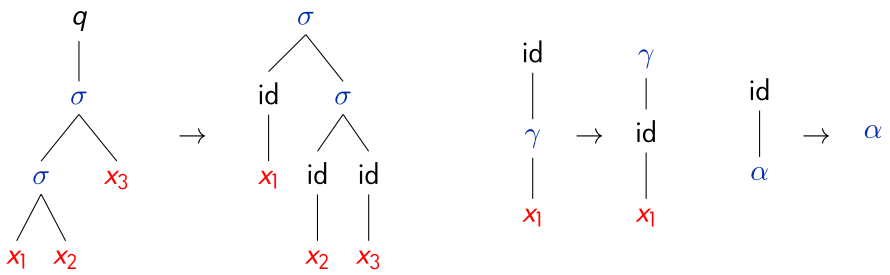
Extended Top-down Tree Transducer

Example rules



Extended Top-down Tree Transducer

Example rules



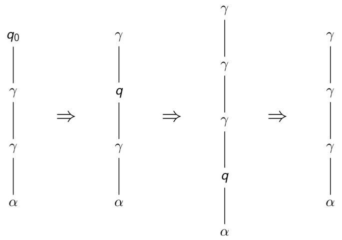
- Rotation via linear transducer (linear, nondeleting defined as before)
- Replace deterministic by functional
  (every state can translate every input tree in at most one manner)
- Replace total by total
  (every state can translate every input tree in at least one manner)

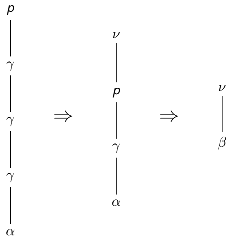
Extended Top-down Tree Transducer

General composition strategy

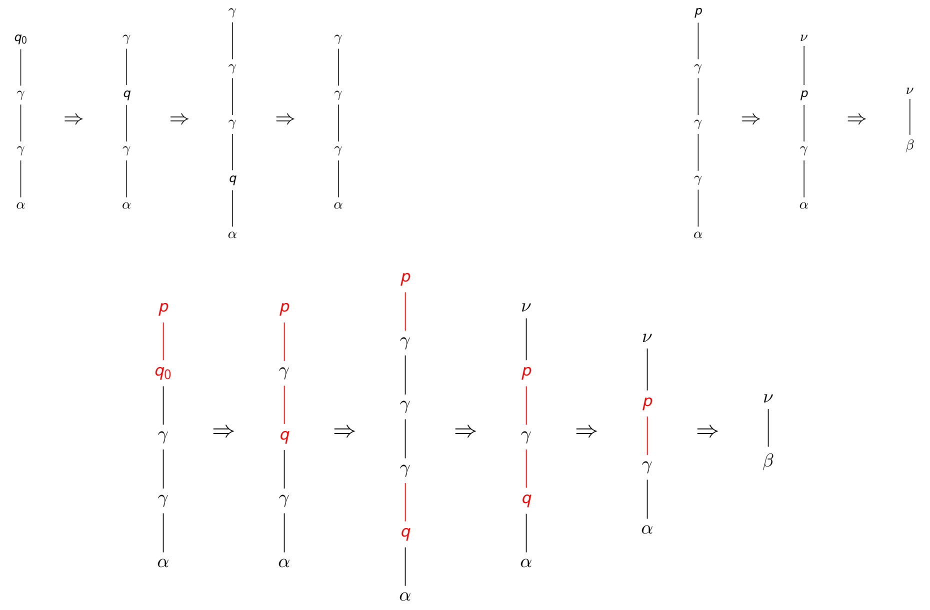
- Still works on derivation level
- But fails on rule level
  (potentially cannot process full intermediate tree fragment)

General Composition Strategy





General Composition Strategy



Composition Results

Classical transducers (unextended)

Case	1st transducer	2nd transducer
(a)		linear and nondeleting
(b)	total	linear
(c)	functional	nondeleting
(d)	functional and total	

- Mixed composition results mostly unknown for extended transducers
- Most results require 1 transducer to be classical (unextended)

Composition Closure

- Composition closure level is smallest number of transducers needed to simulate any composition chain
- Class closed under composition $\iff$ Level $= 1$

Classical transducers

Case	Level	Remarks
linear, nondel., nonerasing	1	closed
linear, nondel.	1	closed
functional, total	1	closed
linear	2	regular look-ahead obtained at level 2
linear with regular look-ahead	1	closed

Composition Closure

Composition closures [Engelfriet, Fülöp, $\sim$ 2017]

Case	Classical	Extended
linear, nondeleting, and nonerasing	1	

Composition Closure

Composition closures [Engelfriet, Fülöp, $\sim$ 2017]

Case	Classical	Extended
linear, nondeleting, and nonerasing	1	2
linear and nondeleting	1	

Composition Closure

Composition closures [Engelfriet, Fülöp, $\sim$ 2017]

Case	Classical	Extended
linear, nondeleting, and nonerasing	1	2
linear and nondeleting	1	∞
functional and total	1	

Composition Closure

Composition closures [Engelfriet, Fülöp, $\sim$ 2017]

Case	Classical	Extended
linear, nondeleting, and nonerasing	1	2
linear and nondeleting	1	∞
functional and total	1	?
linear	2	

Composition Closure

Composition closures [Engelfriet, Fülöp, $\sim$ 2017]

Case	Classical	Extended
linear, nondeleting, and nonerasing	1	2
linear and nondeleting	1	∞
functional and total	1	?
linear	2	4
linear with regular look-ahead	1	

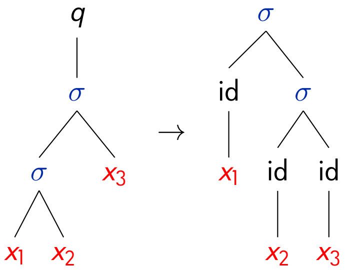
Composition Closure

Composition closures [Engelfriet, Fülöp, $\sim$ 2017]

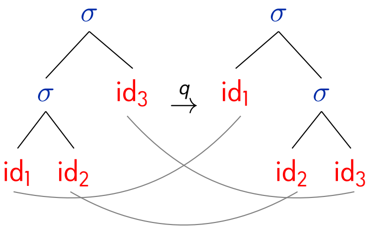
Case	Classical	Extended
linear, nondeleting, and nonerasing	1	2
linear and nondeleting	1	∞
functional and total	1	?
linear	2	4
linear with regular look-ahead	1	3

Slightly Different Viewpoint

Alternative rule representation

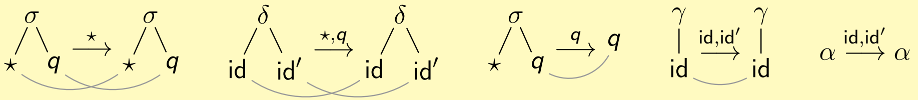


corresponds to



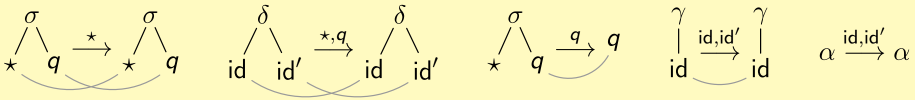
Synchronous Generation

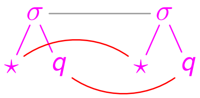
Rules



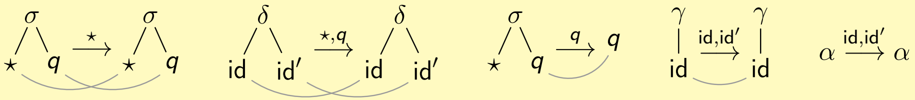
Synchronous Generation

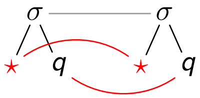
Rules





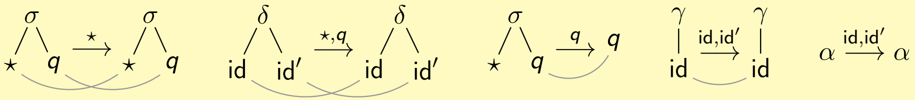
Rules

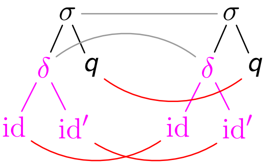




Synchronous Generation

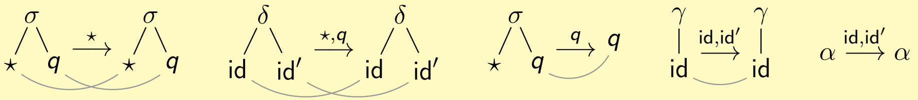
Rules

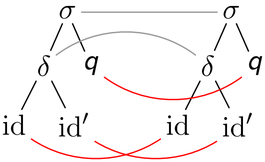




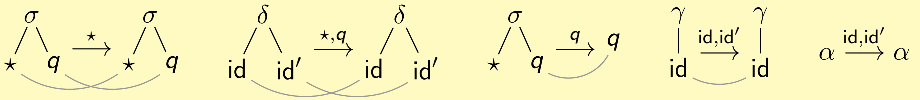
Synchronous Generation

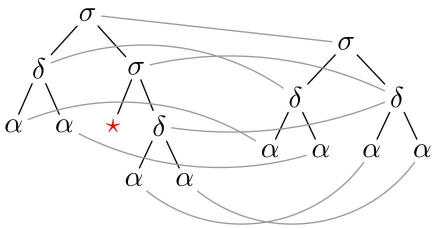
Rules





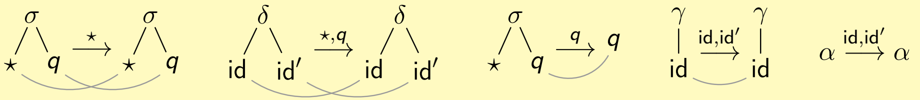
Rules

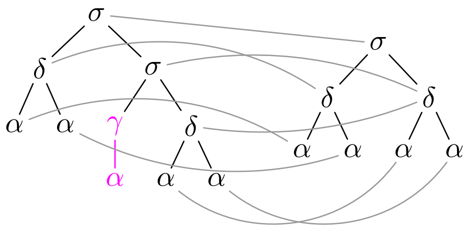

σ ★ q ★ σ ★ q
δ id id′ ★,q δ id id′
σ ★ q q q
γ id id,id′ γ id
α id,id′ α


σ δ σ α α ★ δ α α
σ δ δ α α α α

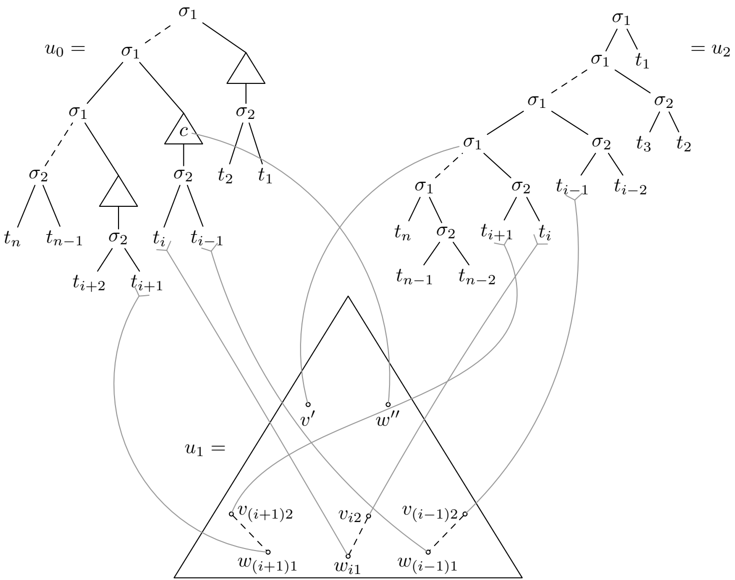
Synchronous Generation

Rules





Composition Closure



Extended Top-down Tree Transducer

- General composition strategy fails on rule level
  due to extended 2nd transducer
- Restrict to extended transducers followed by classical transducers
- Strategy works as expected

Composition results [Engelfriet 1975], [Baker 1979]

Case	Extended 1st transducer	Classical 2nd transducer
(a)		linear and nondeleting
(b)	total	linear
(c)	functional	nondeleting
(d)	functional and total	

Weights

Another extension

- Weights used in practice to resolve nondeterminism
- Each rule is assigned weight
- Weights multiplied along derivation
- Weights of alternatives are added

Typical weights

- Probabilities
- Costs
- Flows
- Profits

Weight Structure

Definition

Commutative semiring $(C, +, \cdot, 0, 1)$ if

- $(C, +, 0)$ and $(C, \cdot, 1)$ commutative monoids
- $\cdot$ distributes over finite (incl. empty) sums

Idempotent if $c + c = c$

Weight Structure

Definition

Commutative semiring $(C, +, \cdot, 0, 1)$ if

- $(C, +, 0)$ and $(C, \cdot, 1)$ commutative monoids
- $\cdot$ distributes over finite (incl. empty) sums

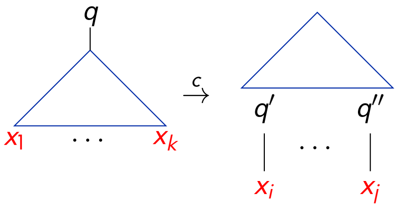
Idempotent if $c + c = c$

Examples

- Boolean semiring $(\{0, 1\}, \max, \min, 0, 1)$ (idempotent)
- semiring $(\mathbb{N}, +, \cdot, 0, 1)$ of non-negative integers
- tropical semiring $(\mathbb{N} \cup \{\infty\}, \min, +, \infty, 0)$ (idempotent)
- any field, ring, etc.

Weighted Extended Top-down Tree Transducer

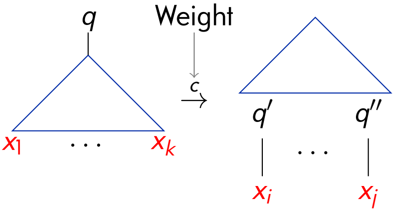
Rule shape [Fülöp, ∼, Vogler 2011]



(Disallow lhs $q(x_1)$)

Weighted Extended Top-down Tree Transducer

Rule shape [Fülöp, ∼, Vogler 2011]



(Disallow lhs q — x_1)

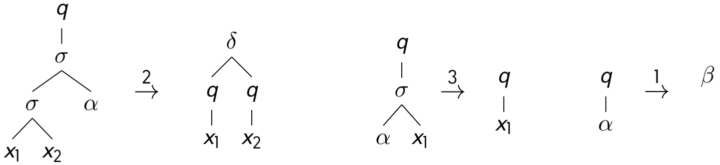
Adjustment of semantics

- Weight of derivation is product of rule weights
  (rule weight taken as often as rule is used)

Adjustment of semantics

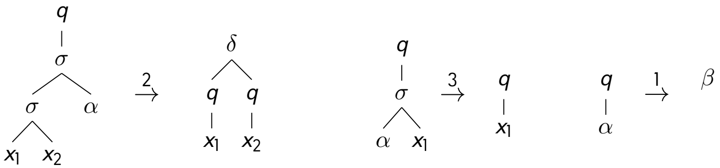
- Weight of derivation is product of rule weights
  (rule weight taken as often as rule is used)
- Weight of translation is sum of all derivations for that translation
  (actually only left-most derivations to normalize rewrite order)

Example rules with integer weights

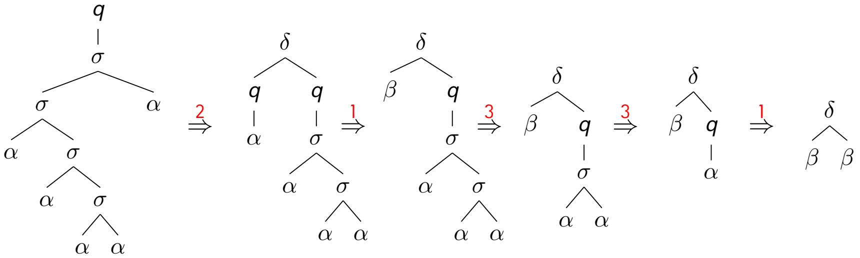


Weighted Extended Top-down Tree Transducer

Example rules with integer weights



Example derivation of weight $2 \cdot 1 \cdot 3 \cdot 3 \cdot 1$



Weighted Composition

Given $\tau_1 \colon T_\Sigma \times T_\Gamma \to C$ and $\tau_2 \colon T_\Gamma \times T_\Delta \to C$

$$(\tau_1 \,;\, \tau_2)(s, u) = \sum_{t \in T_\Gamma} \tau_1(s, t) \cdot \tau_2(t, u)$$

Weighted Composition

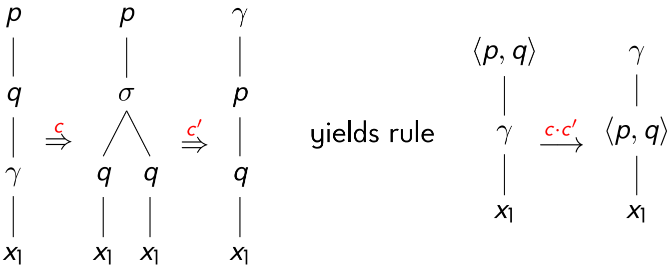
Given $\tau_1 \colon T_\Sigma \times T_\Gamma \to C$ and $\tau_2 \colon T_\Gamma \times T_\Delta \to C$

$$(\tau_1 \,;\, \tau_2)(s, u) = \sum_{t \in T_\Gamma} \tau_1(s, t) \cdot \tau_2(t, u)$$

Notes

- Weighted composition is standard matrix product
  (identify $\tau_1 \in C^{T_\Sigma \times T_\Gamma}$ and $\tau_2 \in C^{T_\Gamma \times T_\Delta}$)
- Distributivity and commutativity allow general composition strategy
  (but derivations have weights now)

Weighted version



Weighted Composition Results

Composition results

Case	Extended 1st transd.	Classical 2nd transd.	Weights
(a)		linear and nondeleting	
(b)	total	linear	
(c)	functional	nondeleting	
(d)	functional and total		

Weighted Composition Results

Composition results

Case	Extended 1st transd.	Classical 2nd transd.	Weights
(a)		linear and nondeleting	✓
(b)	total	linear	
(c)	functional	nondeleting	
(d)	functional and total		

- Case (a) simply works in weighted setting
  [Engelfriet, Fülöp, Vogler 2002], [Lagoutte, ∼ 2011]

Weighted Composition Results

Composition results

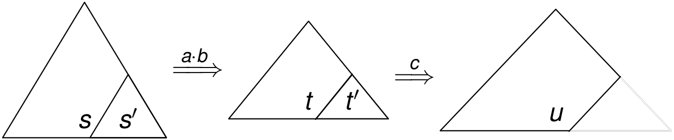
Case	Extended 1st transd.	Classical 2nd transd.	Weights
(a)		linear and nondeleting	✓
(b)	total	linear	?
(c)	functional	nondeleting	
(d)	functional and total		

Weighted Composition Results

Composition results

Case	Extended 1st transd.	Classical 2nd transd.	Weights
(a)		linear and nondeleting	✓
(b)	total	linear	?
(c)	functional	nondeleting	
(d)	functional and total		

Original:



Composition:

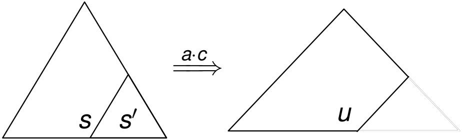
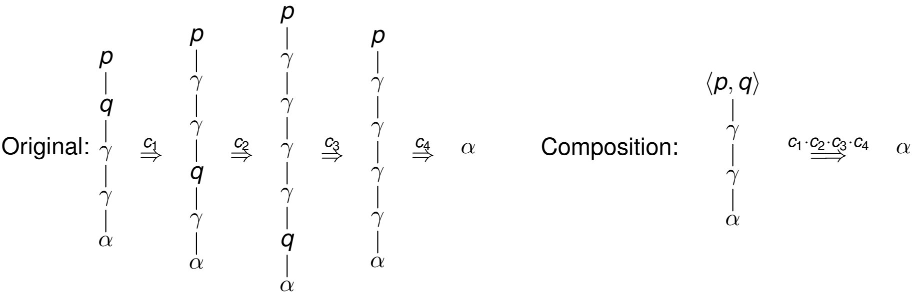


Illustration of rule construction



Weighted Composition Results

Definition

$q \in Q$ constant if there is $c \in C$ such that for every tree $s \in T_\Sigma$

$$\sum_{t \in T_\Gamma} \sum_{\substack{\text{derivation } d \\ \text{from } q(s) \text{ to } t}} \mathrm{wt}(d) = c$$

Weighted Composition Results

Definition

$q \in Q$ constant if there is $c \in C$ such that for every tree $s \in T_\Sigma$

$$\sum_{t \in T_\Gamma} \sum_{\substack{\text{derivation } d \\ \text{from } q(s) \text{ to } t}} \mathrm{wt}(d) = c$$

Examples of 1-constant states

- Total transducer over Boolean semiring (here: constant = total)
- Boolean and total transducer over idempotent semiring
- Functional, total, and Boolean transducer

Weighted Composition Results

Composition results

Case	Extended 1st transd.	Classical 2nd transd.	Weights
(a)		linear and nondeleting	✓
(b)	constant	linear	✓
(c)	functional	nondeleting	
(d)	functional and constant		

Weighted Composition Results

Composition results

Case	Extended 1st transd.	Classical 2nd transd.	Weights
(a)		linear and nondeleting	✓
(b)	constant	linear	✓
(c)	functional	nondeleting	
(d)	functional and constant		

- Case (b) requires "constant" instead of "total"
  [Blattmann, ∼ 2022]

Weighted Composition Results

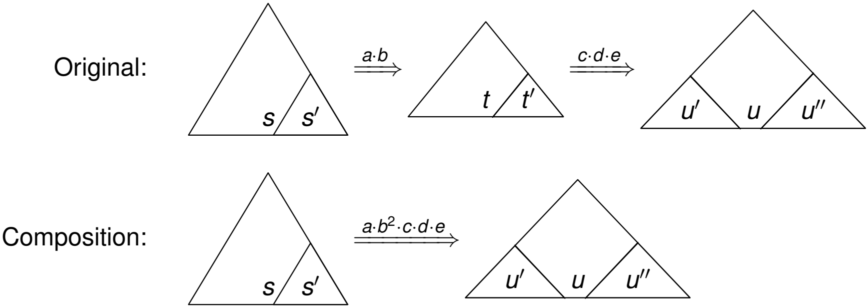
Composition results

Case	Extended 1st transd.	Classical 2nd transd.	Weights
(a)		linear and nondeleting	✓
(b)	constant	linear	✓
(c)	functional	nondeleting	?
(d)	functional and constant		

Weighted Composition Results

Composition results

Case	Extended 1st transd.	Classical 2nd transd.	Weights
(a)		linear and nondeleting	✓
(b)	constant	linear	✓
(c)	functional	nondeleting	?
(d)	functional and constant		



Weighted Composition Results

Composition results

Case	Extended 1st transd.	Classical 2nd transd.	Weights
(a)		linear and nondeleting	✓
(b)	constant	linear	✓
(c)	functional	nondeleting	✓
(d)	functional and constant		✓

Weighted Composition Results

Composition results

Case	Extended 1st transd.	Classical 2nd transd.	Weights
(a)		linear and nondeleting	✓
(b)	constant	linear	✓
(c)	functional	nondeleting	✓
(d)	functional and constant		✓

- Case (c) requires modification to avoid weight production for copies [$\sim$ 2023]
- Case (d) is combination of (b) and (c)

Weighted Composition Results

Composition results

Case	Extended 1st transd.	Classical 2nd transd.	Weights
(a)		linear and nondeleting	✓
(b)	constant	linear	✓
(c)	functional	nondeleting	✓
(d)	functional and constant		✓

- Case (c) requires modification to avoid weight production for copies [$\sim$ 2023]
- Case (d) is combination of (b) and (c)

Thank you for your attention!