The Power of Tree Series Transducers

Andreas Maletti[1]

Technische Universität Dresden
Fakultät Informatik

June 15, 2006

[1]Research funded by German Research Foundation (DFG GK 334)

Machine Translation

Overview

- Subfield of computational linguistics
- Automatic translation shall assist (human) translator
- Offer several (likely) alternatives

History

- 1954: High prospects and expectations after Georgetown Experiment
- 1966: "Perfect translation" failed (ALPAC report)
- 1993: Statistical machine translation system [Brown et al 93]

Machine Translation

Problem

Translate text of language X into grammatical text of language Y.

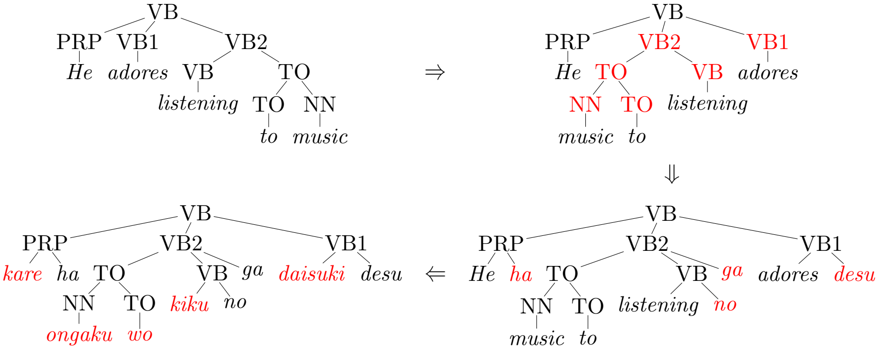
1. Preserve meaning
2. Preserve connotation
3. Preserve style

Relaxed Problem

Transform text of language X into text of language Y such that

1. the result is grammatical
2. expert for X and Y can discern original sentence

Tree-based Model [Yamada, Knight 01]

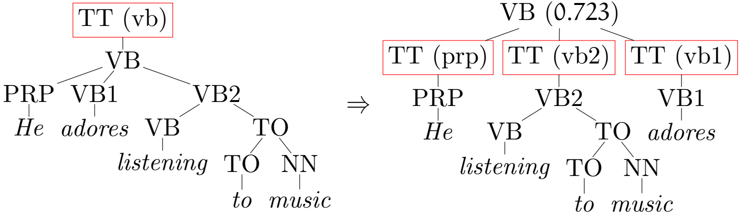


$\boxed{\textit{kare ha ongaku wo kiku no ga daisuki desu}}$

3 phases: (i) Reorder, (ii) Insert, (iii) Translate

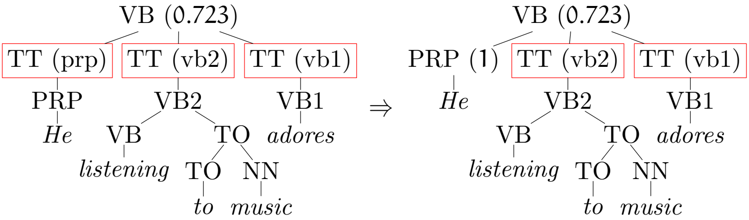
Implementation of Phase (i): Reorder

Implementation by top-down tree series transducer



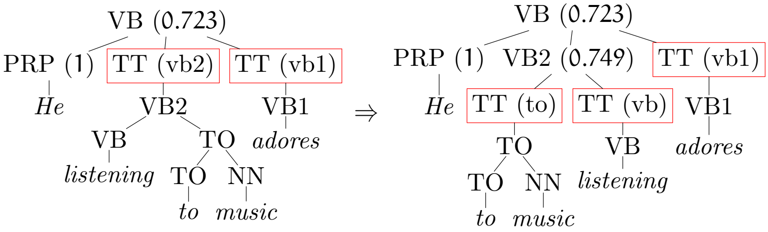
$$\boxed{\mathrm{vb}(\mathrm{VB}(x_1, x_2, x_3)) \stackrel{0.723}{\to} \mathrm{VB}(\mathrm{prp}(x_1), \mathrm{vb2}(x_3), \mathrm{vb1}(x_2))}$$

Implementation of Phase (i): Reorder



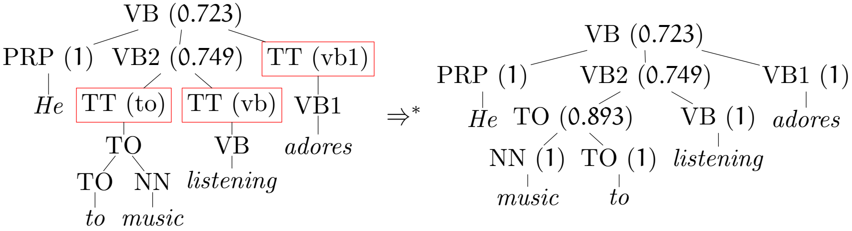
$$\boxed{\mathrm{prp}(\mathrm{PRP}(x_1)) \xrightarrow{1} \mathrm{PRP}(x_1)}$$

Implementation of Phase (i): Reorder

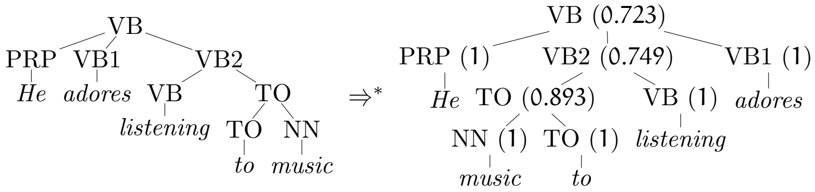


$$\boxed{\mathrm{vb2}(\mathrm{VB2}(x_1, x_2)) \xrightarrow{0.749} \mathrm{VB2}(\mathrm{to}(x_2), \mathrm{vb}(x_1))}$$

Implementation of Phase (i): Reorder



Implementation of Phase (i): Reorder



The above reordering has probability:

$$0.723 \cdot 0.749 \cdot 0.893 = 0.484$$

Implementation Details

Rules

Original	Reordered	Probability
PRP VB1 VB2	PRP VB1 VB2	0.074
	PRP VB2 VB1	0.723
	VB1 PRP VB2	0.061
	VB1 VB2 PRP	0.037
	VB2 PRP VB1	0.083
	VB2 VB1 PRP	0.021
VB TO	VB TO	0.251
	TO VB	0.749
TO NN	TO NN	0.107
	NN TO	0.893

Tiburon [May, Knight 06]

Overview

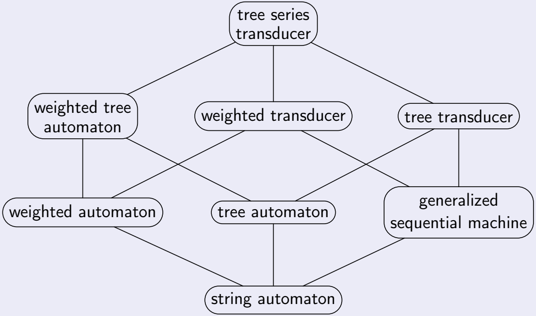
- Implements top-down weighted tree automata and top-down tree series transducers over the probability semiring
- Operations WTA: intersection, weighted determinization, pruning
- Operations TST: application, composition, training

Applications

- Used to implement Yamada-Knight model (custom implementation took > 1 year, implementation in Tiburon 2 days)
- Used to implement Japanese transliteration [Knight, Graehl 98]

Tree Series Transducers

Overview



History

- Introduced in [Kuich 99]
- Extended to full generality in [Engelfriet, Fülöp, Vogler 02]

Semiring

Definition

$(A, +, \cdot, 0, 1)$ semiring, if

- $(A, +, 0)$ commutative monoid
- $(A, \cdot, 1)$ monoid
- $\cdot$ distributes (both sided) over $+$
- 0 is absorbing for $\cdot$ $(a \cdot 0 = 0 = 0 \cdot a)$

Example

- Natural numbers $(\mathbb{N}, +, \cdot, 0, 1)$
- Probabilities $([0, 1], \max, \cdot, 0, 1)$
- Subsets $(\mathfrak{P}(A), \cup, \cap, \emptyset, A)$

Top-down Tree Series Transducer [Engelfriet et al 02]

Definition

Polynomial top-down tree series transducer $(Q, \Sigma, \Delta, \mathcal{A}, I, R)$ where

- Q finite set of states
- Σ and Δ input and output ranked alphabet
- $\mathcal{A} = (A, +, \cdot, 0, 1)$ semiring
- $I \subseteq Q$ set of initial states
- R finite set of rules of the form

$$q(\sigma(x_1, \ldots, x_k)) \xrightarrow{a} t$$

where $t \in T_\Delta(Q(X_k))$

Properties of TST

Definition

$(Q, \Sigma, \Delta, \mathcal{A}, I, R)$ top-down TST

- deterministic, if there is at most one rule with a given left hand side and at most one initial state
- linear, if (for every rule) every variable appears at most once in the right hand side
- nondeleting, if (for every rule) variables that occur in the left hand side also occur in the right hand side

Note

Bottom-up TST process input tree from leaves toward root.

Classes of Transformations

Definition

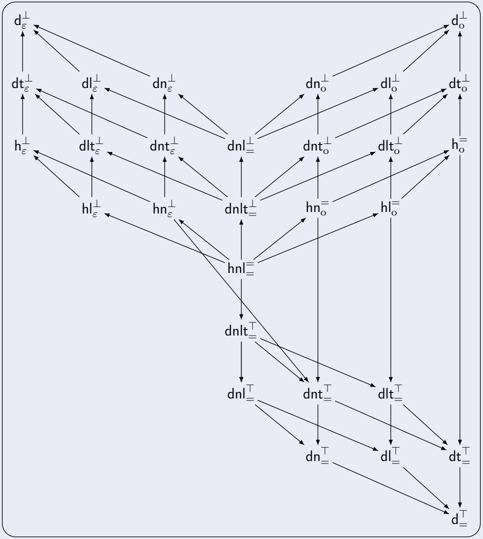
denotation	class of transformations computed by	substitution
$x\text{-TOP}_\varepsilon(\mathcal{A})$	top-down TST with properties x	ε-subst.
$x\text{-TOP}_{\mathrm{o}}(\mathcal{A})$	top-down TST with properties x	o-subst.
$x\text{-BOT}_\varepsilon(\mathcal{A})$	bottom-up TST with properties x	ε-subst.
$x\text{-BOT}_{\mathrm{o}}(\mathcal{A})$	bottom-up TST with properties x	o-subst.

In diagram:

- $x\text{-TOP}_\omega(\mathcal{A})$ abbreviated to $x_\omega^\top$
- $x\text{-BOT}_\omega(\mathcal{A})$ abbreviated to $x_\omega^\perp$

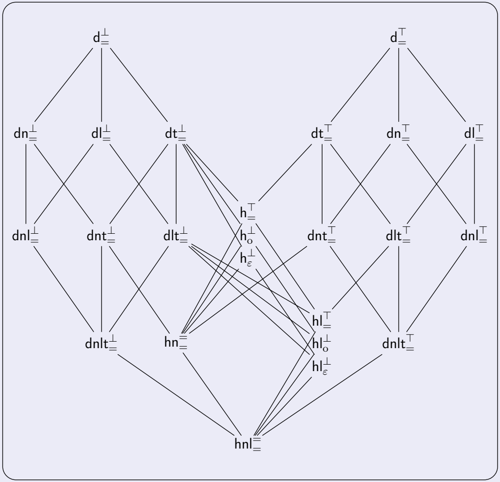
Hasse Diagram for Deterministic TST

Probability Semiring and Semiring of Natural Numbers



Hasse Diagram for Deterministic TST

Semiring of Subsets



Composition of Transformations

Definition

Let

- $\varphi\colon T_\Sigma \times T_\Delta \to A$
- $\psi\colon T_\Delta \times T_\Gamma \to A$

Composition of φ and ψ

$$(\varphi\,;\psi)\colon T_\Sigma \times T_\Gamma \to A$$

$$(t, v) \mapsto \sum_{u \in T_\Delta} \varphi(t, u) \cdot \psi(u, v)$$

Composition Results

Theorem (see [Kuich 99] and [Engelfriet et al 02])

$\mathcal{A}$ *commutative semiring*

$$\text{nlp-BOT}(\mathcal{A}) \,;\, \text{p-BOT}(\mathcal{A}) = \text{p-BOT}(\mathcal{A})$$

$$\text{p-BOT}(\mathcal{A}) \,;\, \text{bdth-BOT}(\mathcal{A}) = \text{p-BOT}(\mathcal{A})$$

Theorem

$\mathcal{A}$ *commutative semiring*

$$\text{lp-BOT}(\mathcal{A}) \,;\, \text{p-BOT}(\mathcal{A}) = \text{p-BOT}(\mathcal{A})$$

$$\text{p-BOT}(\mathcal{A}) \,;\, \text{bd-BOT}(\mathcal{A}) = \text{p-BOT}(\mathcal{A})$$

$$\text{bdt-TOP}(\mathcal{A}) \,;\, \text{lp-TOP}(\mathcal{A}) \subseteq \text{p-TOP}(\mathcal{A})$$

References

Joost Engelfriet, Zoltán Fülöp, and Heiko Vogler.
Bottom-up and top-down tree series transformations.
J. Autom. Lang. Combin., 7(1):11–70, 2002.

Werner Kuich.
Full abstract families of tree series I.
In *Jewels Are Forever*, pages 145–156. Springer, 1999.

Werner Kuich.
Tree transducers and formal tree series.
Acta Cybernet., 14(1):135–149, 1999.

Jonathan May and Kevin Knight.
Tiburon: A weighted tree automata toolkit.
In *Proc. 11th Int. Conf. Implementation and Application of Automata*, LNCS. Springer, 2006.
to appear.

Publications

Conferences

- Myhill-Nerode theorem for sequential transducers over unique GCD-monoids.
  Proc. 9th Int. Conf. Implementation and Application of Automata
- Relating Tree Series Transducers and Weighted Tree Automata.
  Proc. 8th Int. Conf. Developments in Language Theory
- Compositions of bottom-up tree series transformations.
  Proc. 11th Int. Conf. Automata and Formal Languages
- The power of tree series transducers of type I and II.
  Proc. 9th Int. Conf. Developments in Language Theory
- Hierarchies of tree series transformations—Revisited.
  Proc. 10th Int. Conf. Developments in Language Theory
- The substitution vanishes.
  (with A. Kühnemann)
  Proc. 11th Int. Conf. Algebraic Methodology and Software Technology
- Does o-substitution preserve recognizability?
  Proc. 11th Int. Conf. Implementation and Application of Automata

Publications

Journals

- Relating tree series transducers and weighted tree automata.
  Int. J. of Foundations of Computer Science 16, 2005
- Hasse diagrams for classes of deterministic bottom-up tree-to-tree-series transformations.
  Theoretical Computer Science 339, 2005
- Cut sets as recognizable tree languages.
  (with B. Borchardt, B. Šešelja, A. Tepavčević, H. Vogler)
  Fuzzy Sets and Systems 157, 2006
- Incomparability results for classes of polynomial tree series transformations.
  (with H. Vogler)
  J. Automata, Languages and Combinatorics, to appear, 2006
- Bounds for tree automata with polynomial costs.
  (with B. Borchardt, Zs. Gazdag, Z. Fülöp)
  J. Automata, Languages and Combinatorics, to appear, 2006
- Compositions of tree series transformations.
  Theoretical Computer Science, to appear, 2006