Linking Theorems for Tree Transducers[1]

Zoltán Fülöp[a,*], Andreas Maletti[b]

[a] *Department of Foundations of Computer Science, University of Szeged*
Árpád tér 2, H-6720 Szeged, Hungary
[b] *Institute of Computer Science, Universität Leipzig*
Augustusplatz 10–11, 04109 Leipzig, Germany

1. Introduction

Multi bottom-up tree transducers were originally introduced and studied in [1, 2], albeit under different names. We consider the linear and extended variant, which we call MBOT for short. MBOT have good algorithmic properties [3, 4] and thus they were further developed into a formal model for tree-to-tree translation, which is a sub-discipline in syntax-based statistical machine translation (SMT). An open-source implementation of an SMT system based on shallow MBOT is available [5].

The semantics of our MBOT is defined by means of a derivation relation over sentential forms. We apply synchronous rewriting [6], which means that several parts of the sentential form develop (via the rules) at the same time. The left-hand side of a rule contributes to the input tree and the right-hand side to the output tree of the sentential form. For MBOT, the right-hand side consists of a vector of trees, so it can act simultaneously at several positions in the output tree. The input and output positions that are supposed to be developed in parallel are recorded by active links (v, w), which relate a position v in the input tree to a position w in the output tree. After applying a rule using active links, those used links are disabled. Thus disabled links simply record all links that were active at some point during the derivation. In this way, we preserve all links and can later argue about their structure, which will allow us to prove properties about MBOT. Links are similar to origings of [7]. A dependency computed by an MBOT is a triple which consists of an input tree, an output tree derived from it, and the set of all disabled links of the derivation.

Our first result is that the links in each dependency are organized hierarchically and that the distance between (input and output) link targets is bounded (Theorems 1 and 2). Then we provide generic linking theorems for ε-free MBOT which, given an MBOT that computes a tree relation with particular properties, predict certain natural links that must be present in the set of computed dependencies (Theorems 3 and 4). Theorem 3 concerns arbitrary compositions of ε-free XTOP^{R} (which are ε-free MBOT whose right-hand sides contain at most one tree), whereas Theorem 4 concerns a single ε-free MBOT. In both cases, we assume that the computed tree relation contains a sub-relation that is obtained by plugging trees from a simple, yet infinite tree language into an input-output context pair. Finally, we demonstrate in Section 5 how to apply these linking theorems to show that certain tree relations cannot be computed by any ε-free MBOT or by any composition of ε-free XTOP^{R}.

2. Preliminaries

We use the set $\mathbb{N}$ of all nonnegative integers and the set $\mathbb{N}_+$ of all positive integers. The composition of relations ρ and ρ' is denoted by $\rho\,;\rho'$, and the inverse of the relation ρ is denoted by ρ^{-1}.

*Corresponding author

[1] Research of both authors were supported by the German Research Foundation (DFG) grant MA / 4959 / 1-1 and the German Academic Exchange Service (DAAD) and Hungarian Scholarship Board Office (MÖB) exchange project "Theory and Applications of Automata" (grant 5567). The first author was also supported by the program TÁMOP-424B / 2-11 / 1-2012-0001, grant B2 / 2R / 3350 and by OTKA, grant K 108448.

E. Filiot (Ed.): Proc. Workshop Trends in Tree Automata and Tree Transducers, TTATT 2015.

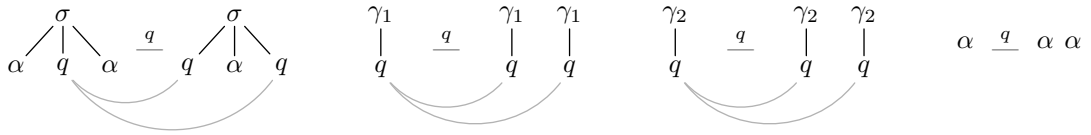

Figure 1: Rules of the MBOT M_{ex}.

The set of all finite words over S is S^*, where $\varepsilon \in S^*$ is the empty word. The concatenation of the words $v, w \in S^*$ is $v.w$ or simply vw. The length of a word $w \in S^*$ is denoted by $|w|$. Given a word (or vector) $w \in \Sigma^*$ and $1 \leq i \leq |w|$, we write w_i for the i^{th} letter in w.

In the following, let Σ be an alphabet and S be a set with $\Sigma \cap S = \emptyset$. The set $T_\Sigma(S)$ of Σ-trees indexed by S is defined as usual, and we let $T_\Sigma = T_\Sigma(\emptyset)$. Let $t \in T_\Sigma(S)$. The set $\mathrm{pos}(t) \subseteq \mathbb{N}_+^*$ of positions of t, the height $\mathrm{ht}(t)$ of t, and the size $|t|$ of t are defined in the standard way. Further, for $w \in \mathrm{pos}(t)$, we denote by $t(w)$ the label of t at w, and by $t|_w$ the w-rooted subtree of t.

Positions are totally ordered by the lexicographic order $\sqsubseteq$ on $\mathbb{N}_+^*$ and partially ordered by the prefix order $\leq$ on $\mathbb{N}_+^*$. Given a finite set $P \subseteq \mathbb{N}_+^*$ of positions, we let $\vec{P} = (w_1, \ldots, w_k)$ be the vector of the positions of P in lexicographic order, where $P = \{w_1, \ldots, w_k\}$ with $w_1 \sqsubset \cdots \sqsubset w_k$. For a sequence $\vec{u} = (u_1, \ldots, u_n)$ of trees and positions $\vec{w} = (w_1, \ldots, w_n)$ of t that are pairwise incomparable with respect to $\leq$, we let $t[\vec{u}]_{\vec{w}}$ denote the tree obtained from t by replacing (in parallel) all subtrees $t|_{w_i}$ at w_i by u_i for all $1 \leq i \leq n$. In the special case $n = 1$, we also use the notation $t[u_1]_{w_1}$.

For every $s \in S$, we let $\mathrm{pos}_s(t) = \{w \in \mathrm{pos}(t) \mid t(w) = s\}$. If $|\mathrm{pos}_s(t)| \leq 1$ for every $s \in S$, then the tree $t \in T_\Sigma(S)$ is linear, and we denote the set of all linear trees of $T_\Sigma(S)$ by $T_\Sigma^{\mathrm{lin}}(S)$. We reserve the sets $X = \{x_i \mid i \in \mathbb{N}_+\}$ and $X_n = \{x_i \mid 1 \leq i \leq n\}$ of variables. A tree $t \in T_\Sigma(X_n)$ is an n-context over Σ if t is linear and all variables of X_n occur in t. The set of all n-contexts over Σ is denoted by $C_\Sigma(X_n)$. Given $c \in C_\Sigma(X_n)$ and $t_1, \ldots, t_n \in T_\Sigma$, we write $c[t_1, \ldots, t_n]$ for $c[\vec{t}]_{\vec{w}}$, where $\vec{t} = (t_1, \ldots, t_n)$ and $\vec{w} = (w_1, \ldots, w_n)$ with $w_i \in \mathrm{pos}_{x_i}(c)$ being the unique position of x_i in c for every $1 \leq i \leq n$.

3. Linear extended multi bottom-up tree transducers

A *multi bottom-up tree transducer* (for short: MBOT) is a tuple $M = (Q, \Sigma, I, R)$ where Q is the alphabet of *states*, $I \subseteq Q$ contains the *initial states*, Σ is the alphabet of *input* and *output symbols* such that $\Sigma \cap Q = \emptyset$, and $R \subseteq T_\Sigma^{\mathrm{lin}}(Q) \times Q \times T_\Sigma(Q)^*$ is the nonempty, finite set of *rules*. We write $\ell \xrightarrow{q} \vec{r}$ for a rule $\langle \ell, q, \vec{r} \rangle \in R$. We require that all states in $\vec{r}$ appear in ℓ for every $\langle \ell, q, \vec{r} \rangle \in R$. If $|\vec{r}| \leq 1$ for all $\ell \xrightarrow{q} \vec{r}$ in R, then M is a (linear) *extended top-down tree transducer with regular look-ahead* [8–10] (for short: XTOP$^{\mathrm{R}}$), and if $|\vec{r}| = 1$ for all $\ell \xrightarrow{q} \vec{r}$ in R, then it is a (linear) *nondeleting* XTOP$^{\mathrm{R}}$ (for short: n-XTOP). Finally, it is *ε-free* if $\ell \notin Q$ for all $\ell \xrightarrow{q} \vec{r}$ in R. Each rule $\ell \xrightarrow{q} \varepsilon$ is a *look-ahead rule* because it can be used to check whether an input subtree belongs to a certain regular tree language [11]. For the remaining discussion, let $M = (Q, \Sigma, I, R)$ be an MBOT.

An example is the ε-free MBOT $M_{\mathrm{ex}} = (\{q\}, \Sigma, \{q\}, R)$ with $\Sigma = \{\sigma, \gamma_1, \gamma_2, \alpha\}$ and the set R of rules containing $\sigma(\alpha, q, \alpha) \xrightarrow{q} \sigma(q, \alpha, q)$, $\gamma_1(q) \xrightarrow{q} \gamma_1(q) \,.\, \gamma_1(q)$, $\gamma_2(q) \xrightarrow{q} \gamma_2(q) \,.\, \gamma_2(q)$, and $\alpha \xrightarrow{q} \alpha \,.\, \alpha$ (see Figure 1).

A *link* is just an element $(v, w) \in \mathbb{N}_+^* \times \mathbb{N}_+^*$. A *sentential form over Q and Σ* is a tuple $\langle \xi, A, D, \zeta \rangle$, where $\xi, \zeta \in T_\Sigma(Q)$ and $A, D \subseteq \mathrm{pos}(\xi) \times \mathrm{pos}(\zeta)$. Elements in A and D are called *active* and *disabled* links, respectively. We denote by $\mathcal{SF}(Q, \Sigma)$ the set of all sentential forms over Q and Σ. The *link structure* $\mathrm{links}_{v,\vec{w}}(\ell \xrightarrow{q} \vec{r})$ of the rule $\ell \xrightarrow{q} \vec{r} \in R$ for positions v and $\vec{w} = (w_1, \ldots, w_{|\vec{r}|})$ with $v, w_1, \ldots, w_{|\vec{r}|} \in \mathbb{N}_+^*$ is

$$\mathrm{links}_{v,\vec{w}}(\ell \xrightarrow{q} \vec{r}) = \bigcup_{p \in Q} \bigcup_{i=1}^{|\vec{r}|} \{(vv', w_i w') \mid v' \in \mathrm{pos}_p(\ell),\, w' \in \mathrm{pos}_p(r_i)\} \ .$$

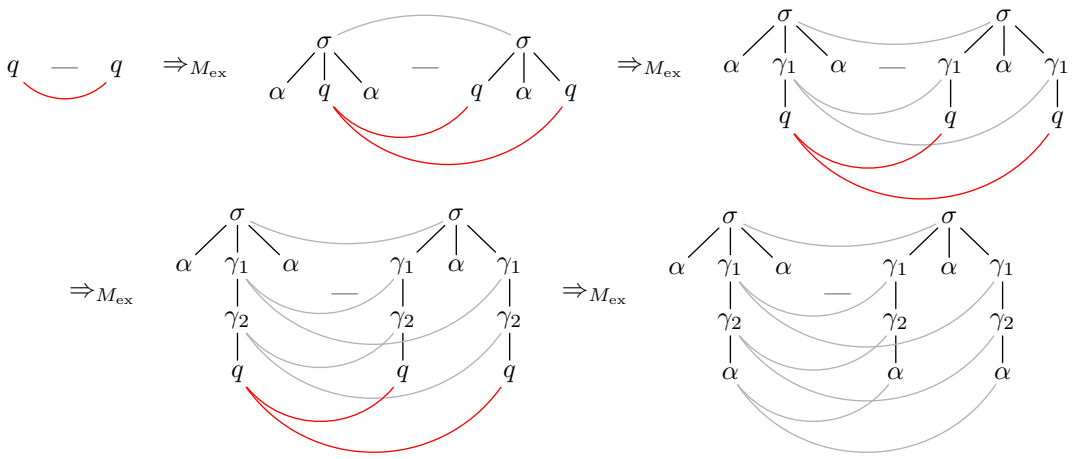

Figure 2: A derivation of the MBOT M_{ex}. The active links are clearly marked, whereas disabled links are shown in light gray.

For the left-most rule ρ presented in Figure 1 and the positions $v = 1.2$ and $\vec{w} = (2)$ we obtain the link structure $\text{links}_{v,\vec{w}}(\rho) = \{(1.2.2, 2.1), (1.2.2, 2.3)\}$. Figure 1 shows the links of $\text{links}_{\varepsilon,(\varepsilon,\ldots,\varepsilon)}(\rho)$ as splines for each $\rho \in R$ of M_{ex}.

Given $\langle \xi, A, D, \zeta\rangle, \langle \xi', A', D', \zeta'\rangle \in \mathcal{SF}(Q, \Sigma)$, we write $\langle \xi, A, D, \zeta\rangle \Rightarrow_M \langle \xi', A', D', \zeta'\rangle$ if there exist a rule $\ell \xrightarrow{q} \vec{r} \in R$, an input position $v \in \text{pos}_q(\xi)$, and actively linked output positions $\vec{w} = A(\vec{v})$ such that (i) $|\vec{r}| = |\vec{w}|$, $\xi' = \xi[\ell]_v$, and $\zeta' = \zeta[\vec{r}]_{\vec{w}}$, and (ii) $D' = D \cup L$ and $A' = (A \setminus L) \cup \text{links}_{v,\vec{w}}(\ell \xrightarrow{q} \vec{r})$ with $L = \{(v, w) \mid w \in A(v)\}$. The *set* $\mathcal{SF}(M)$ *of sentential forms computed by* M is

$$\mathcal{SF}(M) = \{\langle \xi, A, D, \zeta\rangle \in \mathcal{SF}(Q, \Sigma) \mid \exists q \in I \colon \langle q, \{(\varepsilon, \varepsilon)\}, \emptyset, q\rangle \Rightarrow_M^* \langle \xi, A, D, \zeta\rangle\} \ ,$$

and the *set* $\mathcal{D}(M)$ *of dependencies computed by* M is $\mathcal{D}(M) = \{\langle t, D, u\rangle \mid t, u \in T_\Sigma, \langle t, \emptyset, D, u\rangle \in \mathcal{SF}(M)\}$. Finally, the *tree relation computed by* M is $M = \{\langle t, u\rangle \mid \langle t, D, u\rangle \in \mathcal{D}(M)\}$.

A short derivation using the MBOT M_{ex} is shown in Figure 2. It results in the dependency $\big\langle t\,,\, \{(\varepsilon,\varepsilon), (2,1), (2,3), (2.1,1.1), (2.1,3.1), (2.1.1,1.1.1), (2.1.1,3.1.1)\}\,,\, u\big\rangle$, where $t = \sigma(\alpha, \gamma_1(\gamma_2(\alpha)), \alpha)$ and $u = \sigma(\gamma_1(\gamma_2(\alpha)), \alpha, \gamma_1(\gamma_2(\alpha)))$.

Next, we introduce some important properties for sets of links, sentential forms, and the set of dependencies computed by an MBOT (see [12]). A set $L \subseteq \mathbb{N}_+^* \times \mathbb{N}_+^*$ of links is (i) *input hierarchical* if $v_1 < v_2$ implies both $w_2 \not< w_1$ and that there exists $(v_1, w_1') \in L$ with $w_1' \leq w_2$, and (ii) *strictly input hierarchical* if $v_1 < v_2$ implies $w_1 \leq w_2$ and $v_1 = v_2$ implies that w_1 and w_2 are comparable with respect to $\leq$, for all $(v_1, w_1), (v_2, w_2) \in L$. A sentential form $\langle \xi, A, D, \zeta\rangle \in \mathcal{SF}(Q, \Sigma)$ is (strictly) input hierarchical whenever $A \cup D$ is. Finally, $\mathcal{D}(M)$ has those properties if for each $\langle t, D, u\rangle \in \mathcal{D}(M)$ the corresponding sentential form $\langle t, \emptyset, D, u\rangle$ has them [i.e., D has them]. The property *(strictly) output hierarchical* can be defined by requiring the corresponding input-side property for the inverted set L^{-1} of links, the inverted sentential form $\langle \zeta, A^{-1}, D^{-1}, \xi\rangle$, and the set $\mathcal{D}(M)^{-1} = \{\langle u, D^{-1}, t\rangle \mid \langle t, D, u\rangle \in \mathcal{D}(M)\}$.

The links L illustrated in the last derivation step of Figure 2 are input hierarchical. They are not strictly input hierarchical because $(2,1), (2.1, 3.1) \in L$ violates the stricter condition. However, L is strictly output hierarchical.

Theorem 1 (see [12, Lm. 22]) *Let M be an* MBOT. *(i) The set $\mathcal{D}(M)$ is input hierarchical and strictly output hierarchical. (ii) If M is an* XTOP$^{\text{R}}$*, then $\mathcal{D}(M)$ is also strictly input hierarchical.* □

Let $b \in \mathbb{N}$. A sentential form $\langle \xi, A, D, \zeta\rangle \in \mathcal{SF}(Q, \Sigma)$ has (i) *link distance b in the input* if for all links $(v_1, w_1), (v_1 v', w_2) \in A \cup D$ with $|v'| > b$ there exists a link $(v_1 v, w_3) \in A \cup D$ such that $v < v'$ and $1 \leq |v| \leq b$, and (ii) *strict link distance b in the input* if for all positions $v_1, v_1 v' \in \text{pos}(\xi)$ with $|v'| > b$ there exists a link $(v_1 v, w_3) \in A \cup D$ such that $v < v'$ and $1 \leq |v| \leq b$. The set $\mathcal{D}(M)$ of dependencies has those properties

if for each $\langle t, D, u\rangle \in \mathcal{D}(M)$ the corresponding sentential form $\langle t, \emptyset, D, u\rangle$ has them. Moreover, $\mathcal{D}(M)$ is *(strictly) link-distance bounded in the input* if there exists an integer $b \in \mathbb{N}$ such that it has (strict) link distance b in the input. A sentential form $\langle \xi, A, D, \zeta\rangle$ and $\mathcal{D}(M)$ have *(strict) link distance b in the output* if $\langle \zeta, A^{-1}, D^{-1}, \xi\rangle$ and $\mathcal{D}(M)^{-1}$ have (strict) link distance b in the input, respectively.

Theorem 2 *Let M be an* MBOT. *(i) The set $\mathcal{D}(M)$ is link-distance bounded in the input and strictly link-distance bounded in the output. (ii) If M is an n-*XTOP*, then $\mathcal{D}(M)$ is also strictly link-distance bounded in the input.* □

4. Linking theorems

Our linking theorems establish the existence of certain interrelated links, which are forced simply by a subset of the computed tree relation. We need the following utility definitions. A tree $t \in T_\Sigma$ is a *chain* (or unary tree) if $\mathrm{pos}(t) \subseteq \{1\}^*$, and t is a *binary* tree if $\mathrm{pos}(t) \subseteq \{1,2\}^*$. A tree language $T \subseteq T_\Sigma$ is (i) *unary shape-complete* if for every chain $t \in T_\Sigma$ there exists a tree $t' \in T$ with $\mathrm{pos}(t') = \mathrm{pos}(t)$, and (ii) *binary shape-complete* if for every binary tree $t \in T_\Sigma$ there exists a tree $t' \in T$ with $\mathrm{pos}(t') = \mathrm{pos}(t)$.

We now start with a linking theorem for the composition of arbitrarily many ε-free XTOP$^{\mathrm{R}}$. This theorem is only applicable to tree relations, which contain a sub-relation that is obtained with the help of an input and an output context into which we can plug trees from a unary shape-complete tree language. If such a tree relation τ is computed by a composition $\tau = M_1 ; \cdots ; M_k$ of ε-free XTOP$^{\mathrm{R}}$ $M_1, \ldots, M_k$, then we can deduce a dependency and the natural links relating the corresponding subtrees of the contexts.

Theorem 3 *Let $k, n \in \mathbb{N}_+$ and $M_1, \ldots, M_k$ be ε-free* XTOP$^{\mathrm{R}}$ *over Σ such that*

$$\{\langle c[t_1, \ldots, t_n]\,,\, c'[t_1, \ldots, t_n]\rangle \mid t_1 \in T_1, \ldots, t_n \in T_n\} \subseteq M_1 ; \cdots ; M_k$$

for some $c, c' \in C_\Sigma(X_n)$ and unary shape-complete tree languages $T_1, \ldots, T_n \subseteq T_\Sigma$. Then there exist trees $t_1 \in T_1, \ldots, t_n \in T_n$, dependencies $\langle u_0, D_1, u_1\rangle \in \mathcal{D}(M_1), \ldots, \langle u_{k-1}, D_k, u_k\rangle \in \mathcal{D}(M_k)$ with $u_0 = c[t_1, \ldots, t_n]$ and $u_k = c'[t_1, \ldots, t_n]$, and a link $(v_{ji}, w_{ji}) \in D_i$ for each $1 \le i \le k$ and $1 \le j \le n$ such that (i) $\mathrm{pos}_{x_j}(c') \le w_{jk}$ for all $1 \le j \le n$, (ii) $v_{ji} \le w_{j(i-1)}$ for all $2 \le i \le k$ and $1 \le j \le n$, and (iii) $\mathrm{pos}_{x_j}(c) \le v_{j1}$ for all $1 \le j \le n$. □

We know that ε-free MBOT and several relevant subclasses (different from XTOP$^{\mathrm{R}}$ and its subclasses) are closed under composition [3]. Therefore, our second linking theorem concerns a single ε-free MBOT.

Theorem 4 *Let $n \in \mathbb{N}_+$ and $M = (Q, \Sigma, I, R)$ be an ε-free* MBOT *such that*

$$\{\langle c[t_1, \ldots, t_n]\,,\, c'[t_1, \ldots, t_n]\rangle \mid t_1 \in T_1, \ldots, t_n \in T_n\} \subseteq M$$

for some $c, c' \in C_\Sigma(X_n)$ and binary shape-complete tree languages $T_1, \ldots, T_n \subseteq T_\Sigma$. Then there exist trees $t_1 \in T_1, \ldots, t_n \in T_n$, a dependency $\langle c[t_1, \ldots, t_n], D, c'[t_1, \ldots, t_n]\rangle \in \mathcal{D}(M)$ and a link $(v_j, w_j) \in D$ for every $1 \le j \le n$ such that (i) $\mathrm{pos}_{x_j}(c) \le v_j$ for all $1 \le j \le n$ and (ii) $\mathrm{pos}_{x_j}(c') \le w_j$ for all $1 \le j \le n$. □

5. Applications of the linking theorems

We present some applications of our linking theorems to existing results of the literature. We start with a classical result of [1], which states that the class of tree relations computed by XTOP$^{\mathrm{R}}$ (as well as those computed by n-XTOP) is not closed under composition.

Theorem 5 ([1, Sect. 3.4]) *The class of tree relations computable by ε-free* XTOP$^{\mathrm{R}}$ *(or ε-free n-*XTOP*) is not closed under composition.* □

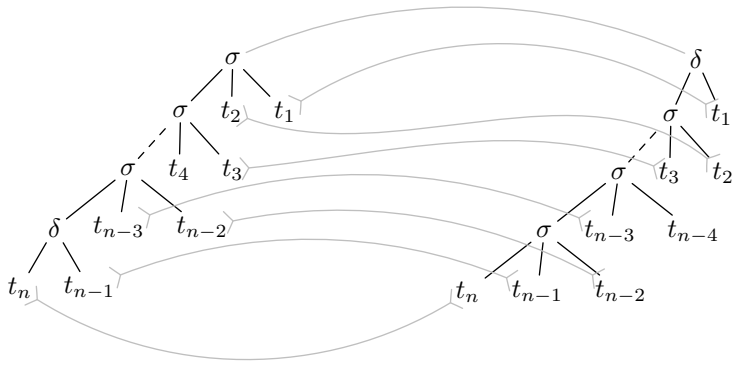

Figure 3: Counterexample relation of [1] with links, which we conclude from Theorem 3, where an inverse arrow head indicates that the link refers to a node (not necessarily the root) inside the subtree that the spline points to.

PROOF Consider the ε-free n-XTOP $M_1 = (Q, \Sigma, \{\star\}, R_1)$ and $M_2 = (Q, \Sigma, \{\star\}, R_2)$ with $Q = \{\star, p, q, r\}$ and $\Sigma = \{\sigma, \delta, \gamma, \alpha\}$, where R_1 contains exactly the rules $\sigma(\star, p, q) \xrightarrow{\star} \delta(\delta(\star, p), q)$, $\delta(p, q) \xrightarrow{\star} \delta(p, q)$, $\gamma(p) \xrightarrow{x} \gamma(p)$, and $\alpha \xrightarrow{x} \alpha$, for all $x \in \{p, q\}$, and R_2 contains exactly the rules $\delta(r, p) \xrightarrow{\star} \delta(r, p)$, $\delta(\delta(r, p), q) \xrightarrow{r} \sigma(r, p, q)$, $\gamma(p) \xrightarrow{x} \gamma(p)$, and $\alpha \xrightarrow{x} \alpha$ for all $x \in \{q, p, r\}$.

Suppose for the sake of a contradiction that there exists an ε-free XTOP$^{\mathrm{R}}$ $M = (Q, \Sigma, I, R)$ that computes $\tau = M_1 \,;\, M_2$. By Theorem 2(i), there is a $b \in \mathbb{N}_+$ such that $\mathcal{D}(M)$ has link distance b in the input. We let $n = 2b + 4$, and as in [1], we select the contexts

$$c = \sigma(\sigma(\cdots \sigma(\delta(x_n, x_{n-1}), x_{n-2}, x_{n-3}) \cdots, x_4, x_3), x_2, x_1)$$
$$c' = \delta(\sigma(\sigma(\cdots \sigma(x_n, x_{n-1}, x_{n-2}) \cdots, x_5, x_4), x_3, x_2), x_1)$$

and the unary shape-complete languages $T_1 = \cdots = T_n = T$, where T is the smallest tree language such that $\alpha \in T$ and $\gamma(t) \in T$ for all $t \in T$. By Theorem 3, there are trees $t_1, \ldots, t_n \in T$, a dependency $\langle c[t_1, \ldots, t_n], D, c'[t_1, \ldots, t_n] \rangle \in \mathcal{D}(M)$, and links $(v_1, w_1), \ldots, (v_n, w_n) \in D$ such that $\mathrm{pos}_{x_j}(c') \leq w_j$ and $\mathrm{pos}_{x_j}(c) \leq v_j$ for all $1 \leq j \leq n$ (see Figure 3). We observe that $(\varepsilon, \varepsilon) \in D$ and $(v_n, w_n) \in D$. By the selection of c, we have $|v_n| > b$, and thus since $\mathcal{D}(M)$ has link distance b in the input, there exists another link $(v, w) \in D$ such that $v < v_n$ and $1 \leq |v| \leq b$. Consequently, $v = 1^m$ for some $1 \leq m \leq b$. Moreover, we observe that $v < v_{2m+2}$ and $v < v_{2m+1}$ because $v < \mathrm{pos}_{x_{2m+2}}(c)$ and $v < \mathrm{pos}_{x_{2m+1}}(c)$. Since D is strictly input hierarchical by Theorem 1(ii), we obtain $w \leq w_{2m+2}$ and $w \leq w_{2m+1}$, which by the shape of c' yields that $w = 1^k$ for some $k \leq m$. However, this also yields that $w < \mathrm{pos}_{x_{2m}}(c') < w_{2m}$. Since D is also strictly output hierarchical by Theorem 1(i), we conclude that $v = 1^m \leq v_{2m}$, which contradicts the shape of c. Thus, we derived the required contradiction and can conclude that such an ε-free XTOP$^{\mathrm{R}}$ cannot exist. ∎

Next we apply Theorem 4 and show that the inverse of abstract topicalization [13] cannot be computed by any ε-free MBOT. A tree language $L \subseteq T_\Sigma$ is *regular* [11] if there exists an MBOT M such that $L = \{t \mid \langle t, u \rangle \in M\}$. A tree relation $\tau \subseteq T_\Sigma \times T_\Sigma$ is *regularity preserving* if $\tau(L) = \{u \mid \langle t, u \rangle \in \tau, t \in L\}$ is regular for every regular tree language $L \subseteq T_\Sigma$.

Theorem 6 ([14, Thm. 8]) *The class of regularity preserving tree relations computable by ε-free* MBOT *is not closed under inverses.* □

PROOF Let $M_{\mathrm{tpc}} = (Q, \Sigma, \{\star\}, R)$ be the ε-free MBOT with $Q = \{\star, p, q, r\}$ and $\Sigma = \{\sigma, \delta, \gamma, \alpha\}$, where R contains exactly the rules $\delta(p, \star) \xrightarrow{\star} \delta(\star, \delta(p, \star))$, $\delta(p, \star) \xrightarrow{\star} \star \,.\, \delta(p, \star)$, $\delta(p, \delta(q, r)) \xrightarrow{\star} r \,.\, \delta(p, q)$, $\sigma(p, q) \xrightarrow{x} \sigma(p, q)$, $\gamma(p) \xrightarrow{x} \gamma(p)$, and $\alpha \xrightarrow{x} \alpha$ for every $x \in \{p, q, r\}$. We can check that M_{tpc} is regularity preserving. The inverse M_{tpc}^{-1}, which is also regularity preserving, is illustrated in Figure 4. We suppose for the sake of a contradiction that there exists an ε-free MBOT $M = (Q, \Sigma, I, R)$ that computes M_{tpc}^{-1}. By Theorem 2(i) there exists a $b \in \mathbb{N}_+$ such that $\mathcal{D}(M)$ has strict link distance b in the output. Moreover, let $n > b + 2$, and we

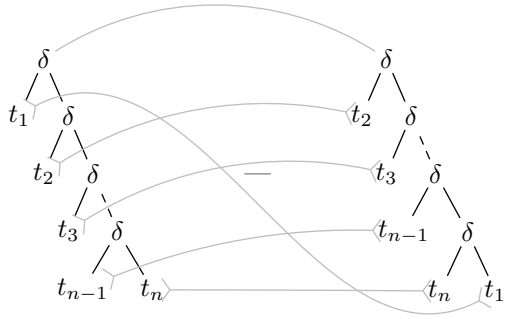

Figure 4: Counterexample relation M_{tpc}^{-1} with links, which we conclude from Theorem 4.

select the contexts

$$c = \delta(x_1, \delta(x_2, \dots \delta(x_{n-1}, x_n) \cdots)) \quad \text{and} \quad c' = \delta(x_2, \delta(x_3, \dots \delta(x_{n-1}, \delta(x_n, x_1)) \cdots))$$

and the binary shape-complete tree languages $T_1 = \cdots = T_n = T$, where T is the smallest tree language such that $\alpha \in T$, $\gamma(t) \in T$ for all trees $t \in T$, and $\sigma(t_1, t_2) \in T$ for all trees $t_1, t_2 \in T$. By Theorem 4 there are trees $t_1, \dots, t_n \in T$, a dependency $\langle c[t_1, \dots, t_n], D, c'[t_1, \dots, t_n] \rangle \in \mathcal{D}(M)$, and a link $(v_j, w_j) \in D$ for every $1 \leq j \leq n$ such that (i) $\text{pos}_{x_j}(c) \leq v_j$ and (ii) $\text{pos}_{x_j}(c') \leq w_j$ for all $1 \leq j \leq n$ (see Figure 4). Based on those links, we can derive a contradiction because there exists a link $(v, w) \in D$ such that $\varepsilon < w < 2^{b+1}$. ■

Finally, we show that the relation M_{tpc} cannot be computed by any composition of ε-free XTOP^{R} (see [14]). The authors believe that a proof approach based on the common fooling technique would be rather difficult (or even hopeless) as we would need to argue over several (at least 2) unknown intermediate trees.

Theorem 7 ([14, Thm. 6]) *The relation M_{tpc} cannot be computed by any chain of ε-free XTOP^{R}.* □

Proof (Sketch.) As before we prove the statement by contradiction. Therefore, we assume that M_{tpc} is computed by a composition of several ε-free XTOP^{R}. By [15, Thm. 11] we know that three ε-free XTOP^{R} suffice, so there are ε-free XTOP^{R} M_1, M_2, and M_3 over Σ such that $M_{\text{tpc}} = M_1 \,;\, M_2 \,;\, M_3$. By Theorem 3 there are links, which can be used to derive a contradiction. ■

References

[1] A. Arnold, M. Dauchet, Morphismes et bimorphismes d'arbres, Theoret. Comput. Sci. 20 (1) (1982) 33–93.

[2] E. Lilin, Propriétés de clôture d'une extension de transducteurs d'arbres déterministes, in: Proc. CAAP, Vol. 112 of LNCS, Springer, 1981, pp. 280–289.

[3] J. Engelfriet, E. Lilin, A. Maletti, Composition and decomposition of extended multi bottom-up tree transducers, Acta Inf. 46 (8) (2009) 561–590.

[4] A. Maletti, An alternative to synchronous tree substitution grammars, J. Natur. Lang. Engrg. 17 (2) (2011) 221–242.

[5] F. Braune, A. Maletti, D. Quernheim, N. Seemann, Shallow local multi bottom-up tree transducers in statistical machine translation, in: Proc. ACL, Association for Computational Linguistics, 2013, pp. 811–821.

[6] D. Chiang, An introduction to synchronous grammars, in: Proc. ACL, Association for Computational Linguistics, 2006, part of a tutorial given with Kevin Knight.

[7] M. Bojanczyk, Transducers with origin information, in: Proc. ICALP, Vol. 8573 of LNCS, Springer, 2014, pp. 26–37.

[8] A. Arnold, M. Dauchet, Bi-transductions de forêts, in: Proc. ICALP, Edinburgh University Press, 1976, pp. 74–86.

[9] J. Engelfriet, Top-down tree transducers with regular look-ahead, Math. Systems Theory 10 (1) (1977) 289–303.

[10] A. Maletti, J. Graehl, M. Hopkins, K. Knight, The power of extended top-down tree transducers, SIAM J. Comput. 39 (2) (2009) 410–430.

[11] F. Gécseg, M. Steinby, Tree languages, in: G. Rozenberg, A. Salomaa (Eds.), Handbook of Formal Languages, Vol. 3, Springer, 1997, Ch. 1, pp. 1–68.

[12] A. Maletti, Tree transformations and dependencies, in: Proc. MOL, Vol. 6878 of LNAI, Springer, 2011, pp. 1–20.

[13] N. Chomsky, The Minimalist Program, Current Studies in Linguistics, MIT Press, 1995.

[14] A. Maletti, The power of regularity-preserving multi bottom-up tree transducers, in: Proc. CIAA, Vol. 8587 of LNCS, Springer, 2014, pp. 278–289.

[15] Z. Fülöp, A. Maletti, Composition closure of ε-free linear extended top-down tree transducers, in: Proc. DLT, Vol. 7907 of LNCS, Springer, 2013, pp. 239–251.