On Defining TALs with Logical Constraints

James Rogers
School of Computer Science
Univ. of Central Florida
jrogers@cs.ucf.edu

In Rogers (1997b) we introduced a new class of models, *three-dimensional tree manifolds* (3-TM), that can serve as both the derived and derivation structures for TAGs in the same way that trees serve as both derived and derivation structures for CFGs. These tree-manifolds are higher-dimensional analogs of trees; in a 3-TM the children of a node form an ordinary (two-dimensional) tree just as in ordinary trees the children of a node form a string. From this point of view the elementary structures of a TAG can be interpreted as labeled local 3-TMs—a root node and its set of children (a pyramidal structure)—analogous to the interpretation of the rewrite rules of a CFG as local trees. Adjunction in TAGs and substitution in CFGs both reduce to a form of concatenation, of local trees in CFGs, of local 3-TMs in TAGs. In Figure 1, for example, the local 3-TMs corresponding to the elementary trees α_1 and β_1 are concatenated to form the 3-TM corresponding to the result of adjoining β_1 into α_1. The two-dimensional yield of this structure is the corresponding derived tree and its one-dimensional yield is the derived string.

This analogy can be extended downward to encompass the regular languages and upward generating the *control language hierarchy* of Vijay-Shanker et al. (1987), Weir (1988), Weir (1992). And it turns out to be quite deep. The ordinary finite-state automata (over strings—the one-dimensional level) accepting the regular languages become, at the two-dimensional level, the tree-automata accepting the recognizable sets of trees. The corresponding automata over 3-TM turn out to accept exactly the sets of tree manifolds that are generated by TAGs (with adjoining constraints) modulo a relaxation of the usual requirement that the root and foot of an auxiliary tree be labeled identically to each other and to the node at which it adjoins. (We refer to these sets as the recognizable sets of three-dimensional tree manifolds.) Moreover, essentially all of the familiar automata-theoretic proofs of properties of regular languages lift directly to automata over tree-manifolds of arbitrary dimension—the dimensionality of the structures is simply a parameter of the proof and plays no essential role.

In Rogers (1998) we exploit this regularity to obtain results analogous to Büchi's characterization of the regular languages in terms of definability in wS1S (the weak monadic second-order theory of the natural numbers with successor) (Büchi, 1960) and Doner's (1970) and Thatcher and Wright's (1968) characterizations of the recognizable sets (of trees) in terms of definability in wSnS (the weak monadic second-order theory of n successor functions—the complete n-branching tree). The recognizable sets of 3-TM are exactly the finite 3-TM definable in the weak monadic second-order theory of the complete n-branching three-dimensional tree manifold, which we refer to as wSnT3. This raises the prospect of defining TALs through the medium of collections of logical constraints expressed in the signature of wSnT3 rather than with explicit TAGs. In this paper, we introduce this approach and begin to explore some of its ramifications in the context of TAGs for natural languages.

Rather than work in wSnT3 directly, we work with an equivalent class of structures that is linguistically more natural. A *Labeled Headed Finite 3-TM* is a structure:

$$\langle T, \triangleleft_3, \triangleleft_2, \triangleleft_1, \bar{\triangleleft}_3, \bar{\triangleleft}_2, \bar{\triangleleft}_1, \triangleleft_3^+, \triangleleft_2^+, \triangleleft_1^+, \mathrm{H}_1, \mathrm{P}_\sigma \rangle_{\sigma\in\Sigma},$$

where T is a rooted, connected, finite subset of the complete n-branching 3-TM (for some n); $\triangleleft_i$ is immediate domination, $\bar{\triangleleft}_i$ is local proper domination (among siblings) and $\triangleleft_i^+$ is global proper domination (inherited), all in the i^{th} dimension;[1] H_1 is the set of *Heads* (exactly one in each string of children—these are underlined in the figures) and P_σ are the labels (each picking out the set of nodes labeled σ, not necessarily mutually exclusive).

We begin by looking at a simple example: assignment of case in XTAG main verb (α_1) and auxiliary verb (β_1) trees. We interpret node names as first-order variables and tree names as monadic second-order variables with, e.g., $\alpha_1(x)$ satisfied iff x is

[1] Domination, in its familiar form in trees, is domination in the second dimension here. Domination in the first dimension is usually known as linear precedence. We will refer to domination in the third dimension as *above*.

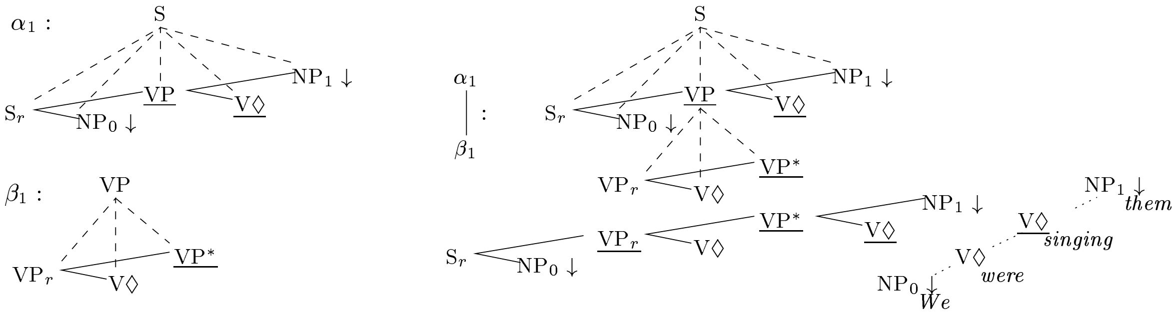

Figure 1: Tree Manifolds

the (3^{rd} -dimensional) root of the local 3-TM corresponding to α_1:

$$
\begin{array}{l}
\alpha_1(s) \leftrightarrow \\
\;(\exists s_r, np_0, vp, v, np_1)[\\
\quad s \lhd_3 s_r \wedge s \lhd_3 np_0 \wedge s \lhd_3 vp \wedge s \lhd_3 v \wedge s \lhd_3 np_1 \wedge \\
\quad \mathrm{Min}_2(s_r) \wedge \mathrm{Max}_2(np_0) \wedge \mathrm{Max}_2(v) \wedge \mathrm{Max}_2(np_1) \wedge \\
\quad s_r \lhd_2 np_0 \wedge s_r \lhd_2 vp \wedge \mathrm{H}_1(vp) \wedge \\
\quad \mathrm{Min}_1(np_0) \wedge np_0 \lhd_1 vp \wedge \mathrm{Max}_1(vp) \wedge \\
\quad vp \lhd_2 v \wedge vp \lhd_2 np_1 \wedge \mathrm{H}_1(v) \wedge \\
\quad \mathrm{Min}_1(v) \wedge v \lhd_1 np_1 \wedge \mathrm{Max}_1(np_1) \wedge \\
\quad \mathrm{Initial}(s) \wedge \mathrm{Anchor}(v) \wedge \mathrm{Subst}(np_0) \wedge \mathrm{Subst}(np_1) \\
\;]
\end{array}
$$

Here Min_i and Max_i pick out minimal (root) and maximal (leaf) nodes wrt the i^{th} dimension—these are defined predicates:

$$\mathrm{Min}_i(x) \equiv \neg(\exists y)[y \lhd_i x].$$

$\mathrm{Initial}(x)$ is true at the root of each local 3-TM encoding an initial tree, $\mathrm{Anchor}(x)$ is true at each anchor node (we will ignore insertion of the lexical items), and $\mathrm{Subst}(x)$ is true at each node marked for substitution—these are labels, in Σ. We require all Subst nodes to have children in the 3^{rd} -dimension and require the set of Initial nodes to be exactly the Subst nodes plus the root of the entire 3-TM:

$$
\begin{array}{c}
(\forall x)[\mathrm{Subst}(x) \rightarrow (\exists y)[x \lhd_3 y]] \\
(\forall x)[\mathrm{Initial}(x) \leftrightarrow (\mathrm{Subst}(x) \vee \mathrm{Min}_3(x))]
\end{array}
$$

Figure 2 shows the distribution of features responsible for case assignment in the XTAG grammar. Following the approach of Rogers (1997a) we interpret the paths occurring in the feature structures decorating the trees as monadic predicates: Σ includes each sequence of features that is a prefix of a path occurring in a feature-structure derivable in the grammar.[2] We will refer to this set of sequences as Feat. Each node is multiply labeled: the feature-structure associated with it is the union of the paths labeling it. In order to capture the distinction between top and bottom feature-structures we will prefix their paths with 't' and 'b', respectively. We can then add to the definition of α_1:

$$\langle \mathrm{t} : \mathrm{case} : \mathrm{acc}\rangle(np_1) \wedge \langle \mathrm{b} : \text{assign-case} : \mathrm{nom}\rangle(v).$$

This encoding of feature-structures gives us a straightforward definition of predicates for path equations as well. For any sequences $w, v \in \mathrm{Feat}$:

$$\langle w = v\rangle(x,y) \equiv \bigwedge_{\substack{w:u \in \mathrm{Feat} \\ \text{or } v:u \in \mathrm{Feat}}} [\langle w : u\rangle(x) \leftrightarrow \langle v : u\rangle(y)].$$

With this we can add the re-entrancy tags:

$$
\begin{array}{l}
\langle \mathrm{b} : \text{assign-case} = \mathrm{t} : \text{assign-case}\rangle(vp, v) \wedge \\
\langle \mathrm{b} : \text{assign-case} = \mathrm{t} : \text{assign-case}\rangle(s_r, vp) \wedge \\
\langle \mathrm{b} : \text{assign-case} = \mathrm{t} : \mathrm{case}\rangle(s_r, np_0) \wedge \\
\langle \mathrm{b} = \mathrm{t}\rangle(s, s_r).
\end{array}
$$

The labeling of the elementary trees can then be interpreted as a collection of constraints on local 3-TM, with the set of structures licensed by the grammar being the set of 3-TM in which every node satisfies one of these collections of constraints. Note that for a 3-TM in which the β_1 3-TM expands the VP node in an α_1 3-TM to be licensed, the VP node must satisfy both the constraints of the α_1 3-TM and the constraints on the root of the β_1 3-TM. Thus the top feature-structure of the VP is unified with the top feature-structure of VP_r and the bottom feature-structure with the bottom feature-structure of the foot VP by simple transitivity of equality. There is no need for additional path equations and no extra-logical mechanisms of any sort; licensing is simply a matter of ordinary model-theoretic satisfaction. To get the (default) unification of top and bottom feature structures of nodes that are not expanded by adjunction we add a single universal principle:

$$(\forall x)[\mathrm{Max}_3(x) \rightarrow \langle t = b\rangle(x,x)].$$

[2] As is typical in FTAG, we assume finite feature-structures.

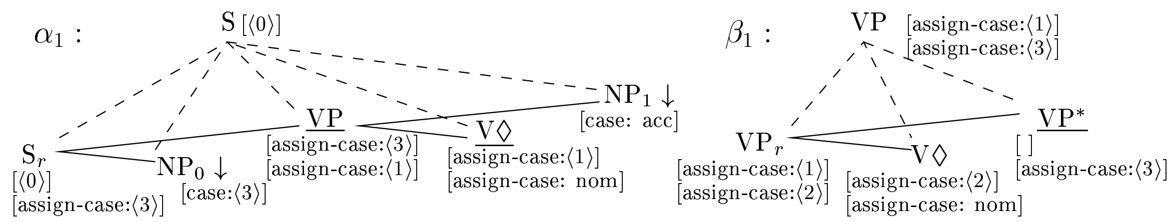

Figure 2: Case assignment in XTAG.

Taken literally, this approach yields little more than a fully declarative restatement of the original grammar. But. in fact, a large proportion of the features decorating elementary trees are there only to facilitate the transport of features through the tree: there is no obvious linguistic motivation for positing that "assign-case" is a feature of VPs or of S. In the language of wSnT3 there is no need for these intermediate "functional" features or even any need to distinguish top and bottom feature structures—we can state directly that the value of the *case* feature of the subject NP, for instance, must agree with the value of the *assign-case* feature of the verb. Of course, what is interesting about this relationship is the effect of adjoined auxiliaries. The TAG analysis includes an *assign-case* feature for the intermediate VP in order to allow auxiliary verbs adjoined at the VP to intercept this relationship by interposing between the VP's top and bottom feature structures. In wSnT3 we obtain the same result from the way in which we identify the relevant verb. For instance, if we take it to be the last adjoined verb[3]—the one most deeply embedded in the third dimension—we can add to the definition of α_1:

$$(\exists x, y)[vp \triangleleft_3^* x \wedge \mathrm{Max}_3(x) \wedge x \overline{\triangleleft}_2 y \wedge \\ \langle\text{assign-case}\rangle(y) \wedge \langle\text{assign-case} = \text{case}\rangle(y, np_0)].$$

In somewhat more linguistically natural terms[4] we might say that a verbal head governs, for the purposes of case assignment, all arguments in its local tree manifold (i.e., the minimal associated structure). Furthermore a verbal head in an auxiliary tree governs all nodes in the structure it adjoins into, as well as all nodes governed by them—effectively each case assigner governs every child of each node properly above it up to the first Initial node:

$$\begin{aligned}&\mathrm{Governs}(x, y) \equiv \\ &\quad \langle\text{assign-case}\rangle(x) \wedge \\ &\quad (\exists z)[z \triangleleft_3^+ x \wedge z \triangleleft_3 y \wedge \\ &\qquad (\forall z')[(z \triangleleft_3^+ z' \wedge z' \triangleleft_3^* x) \rightarrow \neg\mathrm{Initial}(z')]].\end{aligned}$$

Then v assigns case to np_0 iff it governs it and is not, itself, governed by some other case assigner:

$$\begin{aligned}&(\forall x, y)[\\ &\quad (\mathrm{Governs}(x, y) \wedge \langle\text{case}\rangle(y) \wedge \\ &\qquad \neg(\exists z)[\mathrm{Governs}(z, x)]) \rightarrow \\ &\qquad\quad \langle\text{assign-case} = \text{case}\rangle(x, y)].\end{aligned}$$

Alternatively, we could adopt existing accounts based on the more familiar relationships in the two-dimensional projections of the 3-TMs such as traditional GB accounts or Rizzi's (1990) *Relativized Minimality.* All of these are definable in wSnT3 and all, therefore, correspond to some TAG account of case assignment to subjects. The central question, perhaps, is which comes closest to the intuitions informing the existing grammar.

This factoring of a TAG grammar into component linguistic principles is not a new idea. Vijay-Shanker and Schabes's (1992) hierarchical encoding of TAG lexicons using partial descriptions of trees becomes, from this perspective, a matter of classifying the lexicon on the basis of shared properties—every verbal anchor is associated with a subject and the associated structure (see Figure 3):

$$\begin{aligned}&(\forall v)[(\mathrm{Anchor}(v) \wedge \mathrm{Verb}(v)) \rightarrow \\ &\quad (\exists s_r, np_0, vp)[s_r \triangleleft_2 np_0 \wedge s_r \triangleleft_2 vp \wedge vp \triangleleft_2 v \wedge \\ &\qquad \langle\text{case}\rangle(np_0) \wedge \langle\text{assign-case} : \text{nom}\rangle(v) \wedge \cdots]],\end{aligned}$$

transitive verbs, in addition, are associated with an object:

$$\begin{aligned}&(\forall v)[(\mathrm{Anchor}(v) \wedge \mathrm{Verb}(v) \wedge \mathrm{Transitive}(v)) \rightarrow \\ &\qquad (\exists np_1)[v \triangleleft_1 np_1 \wedge \langle\text{case:acc}\rangle(np_1) \wedge \cdots]],\end{aligned}$$

and so on. Note that, since concatenation of 3-TMs does not disturb relationships internal to them, there is no non-monotonicity here (or,rather, the apparent non-monotonicity is an artifact of the yield operation)—there is no need to distinguish top and bottom quasi-nodes, no need for partial trees.

A more obvious connection can be made to Frank's (1992) exploration of universal grammatical principles as interactions of the TAG mechanism with linguistically motivated constraints on the elementary structures. From the current perspective,

[3]This is correct only if the foot nodes have null-adjoining constraints, as is usual.

[4]This is not meant to be a proposal of an analysis of assignment of case in XTAG, only to be an example of the style of analyses that can be supported by this approach.

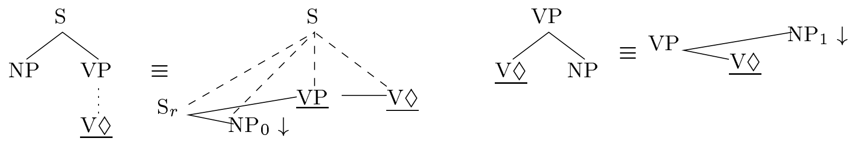

Figure 3: Shared structure as shared properties

these constraints are just properties of the local 3-TMs occurring in well-formed grammatical structures. Here, again, the constraints are not disturbed by the process of building 3-TMs from these local structures—these are properties not just of the elementary structures but of every local 3-TM in all well-formed structures. More interestingly, not all of these constraints are simple properties of the elementary trees, some depend on the derivations. The Specifier Licensing Condition (SLC), for instance, in its basic form, can only be satisfied once an adjunction has taken place. As it turns out, the mechanism employed in capturing this as a condition on the elementary trees is to encode it as a requirement that certain features of the sort we have been calling "functional" are instantiated.[5] Again in this context, in abstracting away from such implementation details, wSnT3 offers a more direct expression of the constraint.

The key feature of this approach is that it isolates the linguistic theory being expressed from the mechanical details of the grammar formalism expressing it—in this respect there is a strong parallel to Mosier's category theoretic approach to HPSG (Mosier, 1997)—*without* losing the restrictions that the formalism imposes. Thus, while the linguistic principles can usually be stated directly, the fact that they must be expressible within the signature of wSnT3 limits them to principles which can be enforced by TAGs. In fact the characterizations of the recognizable sets of 3-TM by definability in wSnT3 and of TAG tree and string languages as the yields of recognizable sets of 3-TM are constructive and when these constructions are carried out many "functional" features of the sort that the logical approach eschews are instantiated in the resulting TAG. This raises the possibility of using the logical definitions not just as an abstract means of discussing the linguistic theory, but also as a sort of higher-level language which can be compiled into TAGs of the familiar sort.[6]

[5] Perhaps coincidentally, these attribute case-assignment to IPs and $\bar{\text{I}}$s in close parallel to the XTAG example we started with.

[6] There are some formidable obstacles to realizing this idea, not the least of which is the fact that the process of compiling wSnT3 formulae into 3-TM automata has, at least potentially, non-elementary complexity. Nonetheless, prior experience at the one- and two-dimensional levels suggests that the process may be feasible even for relatively substantial theories and here we have the knowledge that reasonably compact grammars for similar theories exist (as witnessed by the XTAG grammar). Thus, in some sense, the potential intractability of the theoretical approach may be thought of as a compiler technology issue.

References

J. R. Büchi. 1960. Weak second-order arithmetic and finite automata. *Zeitschrift für mathematische Logik und Grundlagen der Mathematik*, 6.

John Doner. 1970. Tree acceptors and some of their applications. *J. Computer and System Sciences*, 4.

Robert Evan Frank. 1992. *Syntactic Locality and Tree Adjoining Grammar: Grammatical, Acquisition and Processing Perspectives*. Ph.D. thesis, University of Pennsylvania.

M. Andrew Mosier. 1997. Is HPSG featureless or unprincipled. *Linguistics and Philosophy*, 20. (MOL4).

Luigi Rizzi. 1990. *Relativized Minimality*. MIT Press.

James Rogers. 1997a. "Grammarless" phrase structure grammar. *Linguistics and Philosophy*, 20. (MOL4).

James Rogers. 1997b. A unified notion of derived and derivation structures in TAG. In *Proceedings of the Fifth Meeting on Mathematics of Language*, Dagstuhl.

James Rogers. 1998. A descriptive characterization of tree-adjoining languages. In *To Appear: COLING-ACL'98*. Project Note.

J. W. Thatcher and J. B. Wright. 1968. Generalized finite automata theory with an application to a decision problem of second-order logic. *Mathematical Systems Theory*, 2.

K. Vijay-Shanker and Yves Schabes. 1992. Structure sharing in lexicalized tree-adjoining grammars. In *Proceedings COLING'92*.

K. Vijay-Shanker, David J. Weir, and Aravind K. Joshi. 1987. On the progression from context-free to the tree adjoining languages. In A. Manaster-Ramer, editor, *Mathematics of Language*. John Benjamins.

David J. Weir. 1988. *Characterizing Mildly Context-Sensitive Grammar Formalisms*. Ph.D. thesis, University of Pennsylvania.

David J. Weir. 1992. A geometric hierarchy beyond context-free languages. *Theoretical Computer Science*, 104:235–261.