**Modelling Cross-linguistic Variation in Binding Using Synchronous Tree Adjoining Grammar**

Keywords: Syntax, Semantics, Binding, Tree-adjoining-grammar

**Issue**  
Across languages, there are restrictions upon the syntactic distance between bound variables and their antecedents. In formulating an analysis for bound variables in Synchronous Tree Adjoining Grammar (STAG), which derives separate syntactic and semantic tree structures (Shieber and Schabes, 1990), it emerges that these locality constraints can be expressed in terms of a single parametric variation on the derivation of a bound variable.

**Background**  
STAG allows lexical items to be represented by multi-component sets (MCS). For example, QR is simulated by expressing the semantic form of a quantifier in two components: a variable at the quantifier’s interpreted position, and a scope component binding that variable from the quantifier’s scope position. MCS derivations are generally restricted such that all members of the set must combine with the same elementary tree, in this case the predicate. In a derivation tree, where each node represents an elementary tree, structurally recording STAG derivational steps, this restriction is reflected in that MCS members must always be sisters (See Figure 1 for sample derivation). However, Chiang and Scheffler (2008) demonstrates that it is possible to loosen this restriction via delayed tree-locality. Expressed in a derivation tree, delayed tree-locality is licit when both members of an MCS are dominated by a common node. A delay is defined as the set of nodes on the path through the derivation tree between MCS members, excluding the common dominating node, but including the members of the MCS.

**Data**  
As is well known for English, a bound variable pronoun may be used in the same clause as its antecedent (1a), but not as a co-argument of that antecedent (1b).

(1) a. Every girl\(_i\) loves her\(_i\) father. b. * Every girl\(_i\) loves her.

Korean caki, also a bound variable (Han and Storoshenko, 2009), shows no such restriction, and tolerates co-argument binding in addition to less-local relations (2).

(2) a. Motwu\(_i\)-ka caki\(_i\)-uy-emma-lul salang-ha-n-ta.  
   everyone-NOM self-GEN-mother-ACC love-do-PRES-DECL  
   ‘Everyone loves his mother.’

   b. Motwu\(_i\)-ka caki\(_i\)-lul salang-ha-n-ta.  
   everyone-NOM self-ACC love-do-PRES-DECL  
   ‘Everyone loves himself.’

Furthermore, both the English and Korean forms may be bound long distance (across clauses); the key contrast is in how local the binding relation may be.

**Analysis**  
In STAG, bound variables are represented semantically as an MCS: a variable component in the interpreted position, and a binding component which splices into the \(\langle e, t \rangle\) node of the antecedent’s scope component (Storoshenko and Han, 2010). This binding component carries the Binder Index Evaluation Rule of Büring (2005), ensuring the relationship between variable and antecedent (Figure 2). Because the two components must combine with different elementary trees, a delayed tree local derivation is necessary. Nesson and Shieber (2009) proposes that such derivations can be constrained in terms of delay length, herein expressed as \(d\), the cardinality of the delay set. The semantic derivation of (1a) contains three delays: one each for quantificational father of and every girl, and one for the bound variable her. A quantifier MCS combines with a single elementary tree; this local derivation translates to a “trivial” \(d=2\), the smallest possible value. The threshold of delayed locality is where \(d\) reaches 3. Examining the derivation of (1a), for the her MCS, \(d=4\). In the ungrammatical (1b), for the her MCS, \(d=3\) (Figure 3). In the Korean examples, the delays are exactly the same as their English counterparts. Recalling that both languages allow arbitrarily long-distance dependencies, we propose the following constraints on the derivations involving bound variables, based upon this threshold value of \(d=3\):

(3) a. English Bound Variable Pronouns: \(d > 3\)

   b. Korean caki: \(d \geq 3\)

This simple constraint captures the distinction between English (1) and Korean (2).
Conclusion  Being of necessarily delayed tree-locality, all STAG bound variable derivations will have a $d$ of at least 3; considerations of lesser values can be discarded. Thus, we do not expect to find bound variables based upon $d < 3$ or $d \leq 3$. One remaining possibility is a fixed $d = 3$, a variable restricted to co-argument binding, in a structure parallel to (1b). English co-argument reflexives, assuming they have the semantics of bound variables, may fill this role, but in future work, we hope to identify other such forms across languages. Further, by restricting the constraint to the threshold $d$ value of 3, our analysis predicts there to be no variables that need to be at least two (or more) clauses removed from their antecedents, a claim to which we are not presently aware of any counterexamples.

Word count: 747

Figures

**Figure 1:** Sample Semantic STAG Derivation for *Every girl sings.*

**Figure 2:** Semantic Components for Bound Variables

**Figure 3:** Derivation Trees for (1a) and (1b)

References