Abstract

In this paper we present an extension of MCTAGs with Local Shared Derivation (Seddah, 2008) which can handle non-local elliptic coordinations. Based on a model for control verbs that makes use of so-called ghost trees, we show how this extension leads to an analysis of argument cluster coordinations that provides an adequate derivation graph. This is made possible by an original interpretation of the MCTAG derivation tree mixing the views of Kallmeyer (2005) and Weir (1988).

1 Introduction

Elliptic coordinate structures are a challenge for most constituent-based syntactic theories. To model such complex phenomena, many works have argued in favor of factorized syntactic structures (Maxwell and Manning, 1996), while others have argued for distributive structures that include a certain amount of non-lexically realized elements (Beavers and Sag, 2004). Of course, the boundary between those two approaches is not sharp since one can decide to first build a factorized syntactic analysis and then construct a more distributive structure (e.g., logical or functional).

So far, the Combinatorial Categorial Grammar (CCG) framework (Steedman, 2001) is considered as one of the most elegant theories in accounting for coordination. Indeed, the CCG syntactic layer, which is closely tied to an syntax-semantic interface handled in a lexicalized way, permits the coordination of nonstandard constituents that cause a nontrivial challenge for other frameworks. On the other hand, some phenomena such as coordination of unlike categories are still a challenge for theories based on strict atomic category coordination.

In the broader context of ellipsis resolution, Dallymple et al. (1991) propose to consider elided elements as free logical variables resolved using Higher Order Unification as the solving operation. Inspired by this approach and assuming that non-constituent coordination can be analyzed with ellipsis (Beavers and Sag, 2004), we consider elliptic coordination as involving parallel structures where all non-lexically realized syntactic elements must be represented in a derivation structure. This path was also followed by Seddah (2008) who proposed to use the ability of Multi Component TAGs (MCTAGs) (Weir, 1988) to model such a parallelism by including conjunct trees in a same tree set. This simple proposal allows for a straightforward analysis of gapping constructions. The coverage of this account is then extended by introducing links called local shared derivations which, by allowing derivations to be shared across trees of a same set, permit to handle various elliptic coordinate structures in an efficient way. This work showed that, assuming the use of regular operators to handle \( n \)-ary coordinations, a broad range of coordinate structures could be processed using a Tree-Local MCTAG-based formalism named Tree Local MCTAG with Local Shared Derivations. Nevertheless, being tied to the domain of locality of a tree set, the very nature of this mechanism forbids the sharing of derivations between different tree sets, thus preventing it from analyzing non-local elliptic coordinations.

In this paper, we introduce an extension of this model that can handle non-local elliptic coordination — close to unbounded ellipsis (Milward, 1994) — , which can be found in structures involving control verbs and elliptic coordinations.

We also show how our model can cope with argument cluster coordination and why an interpretation of the derivation tree mixing David Weir’s (1988)
original view of MCTAG derivation tree, where each MC set is interpreted as a unique node, and the one introduced by Laura Kallmeyer (2005), where the derivations are the ones from the underlying TAG grammar, is required to yield a derivation tree as close as possible to a proper predicate-argument structure.

2 Standard elliptic coordinate structures

An MCTAG account of many coordinate structures involving ellipsis has been proposed by Seddah (2008). The core idea is to use the extended MCTAG’s domain of locality to enforce a somewhat strict parallelism between coordinate structures.

For example, gapping, as in (1) can be modeled, without any specific operation, by including in a same MC-Set two trees that are identical except for one thing: one is fully lexicalized whereas the other one is anchored by an empty element.

(1) Jean aime Marie et Paul Virginie
John loves Mary and Paul Virginie

Calling this second lexically unrealized tree a ghost tree, the missing anchor can be retrieved simply because the tree it anchors is in the same MC-Set as its ghost tree. In other words, the label of the MC-Set includes the anchor of its fully lexicalized tree. The application of this model to (1) is shown in Figure 1.

Note that this account only requires the expressivity of Tree-Local MCTAGs and that unlike other approaches for gapping in the LTAG framework (Sarkar and Joshi, 1996; Seddah and Sagot, 2006; Lichte and Kallmeyer, 2010), this proposal for gapping does not require any special device or modification of the formalism itself.

In order to model derivations that involve the elision of one syntactic verbal argument as in right node raising cases (RNR) or right subject elision coordinations, the formalism is extended with oriented links, called local shared derivation (local SD), between mandatory derivation site nodes: whenever a derivation is not realized on a given node and assuming that a local SD has been defined between this node and one possible antecedent, a derivation between those nodes is inserted in the derivation structure. Furthermore, if the constraint of having identical tree schema in a tree set (one being fully lexicalized and the other anchored by an empty element) is relaxed, one gets the possibility to give more flexibility to the structure parallelism enforced by this model of gapping. This is what is needed to handle coordination of unlike categories and zeugma constructions (Seddah, 2008).

In the same spirit, by viewing the anchoring process as a regular derivation, and hence allowing local SDs to occur on anchoring derivations as well, one can get a very flexible model allowing for trees, sharing the same tree schema but with different anchors, to be coordinated. Thus, RNRs are simply analyzed in this framework by having two identical tree schema anchored by two different verbs and with one local shared derivation occurring from the N1 node of the right conjunct tree to the N1 of its left counterpart. Such an analysis of RNR for (2) is shown on Figure 2.

(2) Jean fabrique et Marie vend des crêpes,
John makes and Mary sells pancakes

3 MCTAG with Local Shared Derivations

Following Kallmeyer (2005), we define an MCTAG as a tuple $G_{MCTAG} = \langle I, A, N, T, S \rangle$, where $I$ (resp. $A$) is the set of initial (resp. auxiliary) trees, $N$ (resp. $T$) the set of nonterminal (resp. terminal) labels and $S$ the set of elementary MC-Sets. A MC-
**TAG with Local Shared Derivations (MCTAG-LSD)**

$G$ whose underlying MCTAG is $G_{\text{MCTAG}}$ is defined as $G = (I, A, N, T, S, L)$, where $L$ is the set of oriented links between two leaf nodes of two trees in a same MC-Set in $S$.

MCTAG-LSD derivations extend derivations of the underlying MCTAG by allowing for **local shared derivations**, that we shall now define.

Let $\Gamma = \{\gamma_0, \ldots, \gamma_n\}$ be an MC-Set in $S$. Let $L_\Gamma$ be the set of (oriented) links in $\Gamma$, i.e. pairs of the form $(N_L, N_R)$ where $N_L$ and $N_R$ are nodes in two different trees in $\Gamma$. Let us suppose that:

- a tree $\gamma^\prime$ is substituted on a node $N_L$ in a tree $\gamma_i$.
- there exists a node $N_R$ in another tree $\gamma_j$ such that $(N_L, N_R)$ is in $L_\Gamma$.

Then, a local shared derivation can be created as follows:

- a substitution link between $\gamma^\prime$ and $\gamma_j$ is added in the derivation structure; thus, $\gamma^\prime$ has at least two ancestors ($\gamma_i$ and $\gamma_j$) in the derivation structure, which becomes a DAG instead of a tree;
- an initial tree anchored by an empty element is substituted on the node $N_R$.

Note that this also applies for mandatory adjunctions, besides substitutions.

Any MCTAG derivation is a valid MCTAG-LSD derivation. However, local shared derivations allow for performing additional derivation operations. Therefore, the language generated by $G$ strictly contains the language generated by $G_{\text{MCTAG}}$. However, these additional derivations can be simulated in a pure MCTAG fashion, as follows. For a given MCTAG-LSD MC-Set that contains a unique local shared derivation link, we can generate two MCTAG MC-Sets, one that would enforce the substitution by lexicalized trees at both ends of the link, and one that would enforce the substitution of a lexicalized tree at the starting node of the link and the substitution of a **ghost tree** at the other end of the link. This mechanism can be generalized to MC-Sets with more than one local shared derivation. This sketches the proof that the set of languages generated by MCTAG-LSDs is the same as that generated by MCTAGs. Therefore, MCTAG-LSDs and MCTAGs have the same weak generative capacity. Moreover, these considerations still hold while restricting $G_{\text{MCTAG}}$ to be TL-MCTAG. Therefore, TL-MCTAG-LSDs and TL-MCTAGs have the same weak generative power.

In order to cope with very large grammar size, the use of regular operators to factorize out TAG trees has been proposed by (Villemonte de La Clergerie, 2005), and has lead to a drastic reduction of the number of trees in the grammar. The resulting formalism is called **factorized TAGs** and was adapted by Seddah (2008) to the MCTAG-LSD framework in order to handle n-ary coordinations. The idea is to factorize MCTAG-LSD sets that have the same underlying MCTAG set (i.e. they are identical if links are ignored). Indeed, all such MC sets can be merged into one unique tree set associated with the union of all corresponding link sets. However, as with factorized TAGs, we need to add to the resulting tree set a list of constraints, $R$, on the construction of local shared derivations. The result is an extended formalism, called **factorized MCTAG-LSD**, which does not extend the expressive power of MCTAG-LSD but allows for more compact descriptions. Our resulting coordination scheme is shown on Figures 3 and Figure 4.

---

5 Another possibility would be to merge $N_R$ with $N_L$, as for example in (Sarkar and Joshi, 1996). However, this leads to derived DAGs instead of trees.
node through percolation of its feature structure,\(^8\) thus making the underlying unrealized derivation explicit.\(^9\) Figure 6 gives an overview of the process leading to a derivation graph.

4 The case for Unbounded Ellipsis

The problem with this model is its heavy dependence on the domain of locality of a tree set. In fact, if creating a link between two derivation site nodes inside the same tree set is straightforward, things become complicated if the derivations that must be shared involve two nodes from different tree sets. For example, in cases involving a control verb and right-subject ellipsis such as in (3), the subject is shared among the three verbs, although the control verb elementary tree (see Figure 6) cannot be in the same tree set as the others.\(^5\)

\((3)\) Jean\(_i\) ronfle\(\epsilon_i\) espère\(\epsilon_i\) dormir

*John\(_i\) snores and \(\epsilon_i\) hopes \(\epsilon_i\) to sleep*

4.1 Control Verb and MCTAG

Regarding the control verb phenomenon, an LTAG analysis was proposed by Seddah and Gaiffe (2005)\(^7\) involving a complex parsing device, the so-called argumental fusion, and a lexicon based information structure, the control canvas, stating which argument is controlled by the verb (e.g. subject for to hope and object for to forbid). The idea there was to view control verbs as capable of transferring their controlled argument to the trees in which they adjoin by the means of partial derivations, allowing for the creation of a pseudo-derivation between the argument of the control verb tree (i.e. Control Tree) and the embedded verb. This pseudo-derivation accounts for the fact that a syntactic argument of the embedded verb is not realized whereas its morphological features are actually transferred from the Control Tree substitution.

\(^5\)We assume a non-VP coordination analysis of (3).

\(^6\)See this example of feature transfer in French:

\text{Marc_FEM-SG} espère\(\epsilon_i\) être belle

\text{Mary-FEM-SG} hopes \(\epsilon_i\) to be pretty-FEM-SG.

\(^7\)The pure LTAG analysis of French control verbs was initially proposed by Abeillé (1998).

\(^8\)This mismatch between the derivations underlying a derived structure and the real derivation structure is also noted by Kallmeyer (2002) for quantifier and verb interrelations.

\(^9\)Thanks to Timm Lichte for bringing this case to our attention.
Sets for providing an elegant analysis of (3) while remaining in TL-MCTAG-LSD. Nevertheless, nothing prevents us from allowing nominal trees to be characterized in the same way. This allows a (lexically) anchored tree to substitute into a tree of a given MC-Set while one of its ghost trees substitutes into another tree from a different tree set. Thus, it becomes possible to substitute a tree anchored by Jean into the tree anchored by dormir, while its unrealized counterpart will substitute into the argument node of the control verb, therefore allowing the derivation tree displayed in Figure 5a. As one tree is derived into one MC-Set and its ghost tree into another, this analysis falls beyond TL-MCTAG, and benefits from the larger expressivity of NL-MCTAGs.

It shall be noted that having an unrestricted potential number of unrealized ghost trees inside a nominal MC-Set means that a substitution of such a ghost tree can occur in lieu of a shared derivation, thus allowing coindexations of derivation nodes instead of sharing (cf. Figure 5b).

This potential source of ambiguity could be circumvented by stating precedence rules between shared derivations and ghost derivations (i.e. derivation of ghost trees). Nevertheless, such an ambiguity is precisely what is needed to provide an analysis of argument cluster coordination in our framework, as we shall now demonstrate.

5 Argument cluster coordination

Assuming an ellipsis analysis for argument cluster coordination (ACC; (Beavers and Sag, 2004)), sentences such as (4) can be simply analyzed as a case of gapping plus a right subject elision in our framework. This requires an MC-Set α-donne which includes a tree anchored by donner/give and its ghost tree, as depicted in Figure 8.

(4) Jeani donnej une fleur à Marie et εi εj une bague à Paul
John gives Mary a flower and Paul, a ring

However, let us assume an analysis involving a right subject elision and a gapping of the main verb. Then, using the extension of our framework that we defined for handling unbounded ellipsis (section 4), the subject of εj can be obtained in two different ways: (i) via a local shared derivation as sketched in the previous sections (no ghost tree is needed in the MC-Set α-Jean, which contains one unique tree); or (ii) as a ghost tree that belongs to the MC-Set α-Jean.

Note that if we follow Weir’s (1988) original definition of MCTAG derivation, both ways to obtain the subject lead to the same derivation structure. Our own model implies that derivation steps with LSD or involving ghost trees will lead to different structures. This comes from the fact that our model is
based on Kallmeyer’s per-tree interpretation of MC-TAG derivation.

More precisely, Weir’s definition of MCTAG derivation always implies a sharing, whereas Kallmeyer’s own definition leads to two different, possibly co-indexed, nodes. These two possible interpretations of derivation can handle the difference between (i) an elided anchor that refers to the same individual or event as the anchor of the lexicalized tree in the same MC-Set (as Jean in (4)) and (ii) an elided anchor that refers to another (co-indexed) instance of the same class of individuals, or events, (as fleur/flower in (5)).

\[
\begin{align*}
\text{Jean}_i & \text{ donne}_j \text{ une fleur}_k \text{ bleue à Marie et} \\
\text{John}_i & \text{ gives}_j \text{ a blue flower}_k \text{ to Mary and} \\
\text{ε}_i & \text{ ε}_j \text{ une ε}_k \text{ rouge à Paul} \\
\text{ε}_i & \text{ ε}_j \text{ a red (one)_k to Paul}
\end{align*}
\]

Therefore, what we need is a mechanism that can determine whether a given MC-Set denotes a unique event or individual, the latter corresponding to the sharing case or a list of events or individuals that are instances of the same class of events or individuals. Such a mechanism requires more than just syntactic information, typically it needs to rely on an adequate type system.

Let us consider again example (5). Whatever the interpretation of the derivation operations, the derivation runs as follows. Nominal MC-sets α-fleur and α-Jean include ghost trees, whereas the auxiliary trees /β-bleu and /β-rouge have no ghost trees. The auxiliary tree in /β-bleu adjoins to the non-ghost tree in α-fleur while the one in /β-rouge adjoins to the ghost tree in α-fleur. The determiners are treated in the same way. Next, the tree based on the non-ghost tree in α-fleur substitutes in the non-ghost tree in α-donner, whereas the other tree substitutes in the ghost tree in α-donner.12 The gapping and right subject elision are then handled as in Section 2.

Now, let us suppose that we associate the MC-Set α-Jean with a type \(<e>\) and the MC-Set α-fleur with type \(<e, \tau>\). Let us postulate that we use Kallmeyer’s per-tree interpretation for MC-Sets with type \(<e, \tau>\) and Weir’s interpretation for MC-Sets with type \(<e>\), the resulting derivation structure would be exactly the expected predicate-argument structure as shown in Figure 9b and will only require the expressive power of Set Local MCTAGs.

To show how such a structure could be generated, we assumed a rather naive syntax-semantics interface where all elements of a nominal MC-set have the same scope, regardless of their semantic types. That is, as pointed out by an anonymous reviewer, if an NP is right-node-raised, or undergoes a right-subject elision,13 we can have an NP with type \(<e, \tau>\) that leads to a wide scope reading which would imply a single node in the derivation tree. In fact, should we want to distinguish between narrow and wide scope readings, we would need a richer model that could infer scope information from all trees of a MC-set. It would be very interesting to see how a model à la Kallmeyer and Joshi (2003) could be integrated in our framework. In fact, the idea of adding another type of node carrying scope information through the derivation structure seems natural considering the nature of our proposal.

6 Discussion

If syntactic and semantic structures were tied by a strict isomorphism, the TAG derivation tree, with its

---

11 Allowing unlimited adjunction of ghost auxiliary trees would lead to many spurious ambiguities, whereas having modal verbs or adverbs together with their ghost trees in a MC set would certainly be a step toward an elegant treatment of elided modifiers.

12 To avoid spurious ambiguities when ghost trees are substituted, Local Shared Derivations could be used to check that the right ghost tree has been derived wrt to its antecedent.

13 e.g., [Someone from NY], seems to have won the cup and \(\varepsilon\), is likely to win the lottery.
strict encoding of subcategorized arguments, could have been considered as a proper predicate-argument structure. Unfortunately, due to a lack of expressive power, most of the complicated cases of mismatch between syntax and semantics cannot be formalized without breaking the elegance of TAGs’ main property, namely that dealing with elementary trees means dealing with partial dependency structures. Over the last fifteen years, solving this problem has mobilized many teams, and, as noted by (Nesson and Shieber, 2006), led to the emergence of two schools. One focusing on giving more expressive power to the formalism in order to ease either a tight integration between the logical and the syntactic layers (Kallmeyer and Joshi, 1999; Gardent and Kallmeyer, 2003) or a capacity to handle, for instance, free word order languages (Lichte, 2007). The other school focuses either on keeping the syntactic TAG backbone as pure as possible, by designing a new derivation operation to handle coordination (Sarkar and Joshi, 1996) or on carefully designing a syntax–semantic interface built upon TAG derivations (Shieber and Schabes, 1990; Shieber and Nesson, 2007). Our proposal stands in between as we acknowledge that pure TAGs are not powerful enough to carry on simple analysis of complex phenomena while bringing the derivation tree closer to a predicate-argument structure. Recent proposals in the synchronous TAG framework share the same concern. In fact, Shieber and Nesson (2007) use Vector MCTAG (Rambow, 1994), for its ability to underspecify dominance relations and provide the synchronized logical layer with a derivation structure suitable for the analysis of control verbs. However, as we have shown, our solution for control requires a generalization of the mechanism designed for handling elliptic coordination that needs the expressive power of Non Local MCTAGs and tight integration of our proposal with a syntax-semantic interface. This raises two open questions: What generative power do we really need to build appropriate derivation structures? More importantly, where do we want syntax to stop?

7 Conclusion

We have shown how to extend an MCTAG account of coordination with a simple mechanism added on top of its extended domain of locality and which enables the handling of more complex constructions involving control verbs and elliptic coordinations. We have also shown how argument cluster coordinations could be treated in our framework without any special treatment besides the inclusion of a small type inference system if one wants to provide a proper dependency structure. Our work also shows that our treatment of such coordinate constructions needs the expressive power of Non Local MCTAGs to cope with unbounded ellipsis and Set Local MCTAGs for ACC.

References


Figure 9: Sketch of derivations for Argument Cluster Coordination of sentence 5 (John$_i$ gives$_j$ a blue flower$_k$ to Mary and $\varepsilon_i$ $\varepsilon_j$ a red (one)$_k$ to Paul)


