Can the TAG derivation tree represent a semantic graph ? An answer in the light of Meaning-Text Theory.

Marie-Hélène Candito & Sylvain Kahane
TALANA, Université Paris 7 2, place Jussieu, case 7003, 75251 Paris cedex 05
marie-helene.candito@linguist.jussieu.fr sk@ccr.jussieu.fr

Introduction

From the parsing point of view, the derivation tree in TAG [hereafter DT] is seen as the "history" of the derivation but also as a linguistic representation, closer to semantics, that can be the basis of a further analysis.
Because in TAG the elementary trees are lexicalized and localize the predicate-arguments relations, several works have compared the DT to a structure involving dependencies between lexical items (RJ92; RVW95).[1] We agree with these authors that there are divergences between the DT and syntactic dependencies, but we show here that the DT — in the sense of (SS94) — can be viewed as a semantic dependency graph, namely a SemS for Meaning-Text Theory [MTT] (ZM67; M88). This requires the predicate-argument cooccurrence principle and also constraints on the adjunction of predicative auxiliary trees. We briefly introduce the representation levels in MTT before studying the dependencies shown by the DT.[2]

1. Representation levels in MTT

MTT distinguishes between linguistic representations and correspondance rules to go from a representation to another, at an adjacent level. For a written sentence, there are 5 representations, each with a central structure : semantic [SemS], deep and surface syntactic [DSyntS and SSyntS], deep and surface morphological [DMorphS and SMorphS]. At each level, additional structures may supplement the central structure.
A key feature of MTT is that it distinguishes between semantic and syntactic dependencies. The SemS is a graph showing *semantic dependencies* between *semantemes* (= semantic units). The dependencies are numbered to distinguish between the different arguments of a predicative semanteme. An additional structure (the Sem-CommS) indicates communicative features (theme-rheme, focus …). Figure 1 shows an example of SemS for :
(1) The new library owns the book that Peter thinks Mary needs

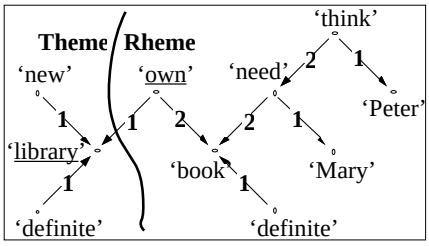

Figure 1 : SemS + Sem-CommS for

The DSyntS (Figure 2) is a dependency tree whose nodes are *generalized lexemes* (= lemma or set of lemmas corresponding to a semantic unit). Its arcs are *deep syntactic dependencies*, that are language independent (6 actancy relations I, II, ...VI, plus ATTR, COORD and APPEND). The SSyntS is a dependency tree showing grammatical relations — language dependent — between lexemes, that may be semantically void. Word order is defined at the deep morphological level.

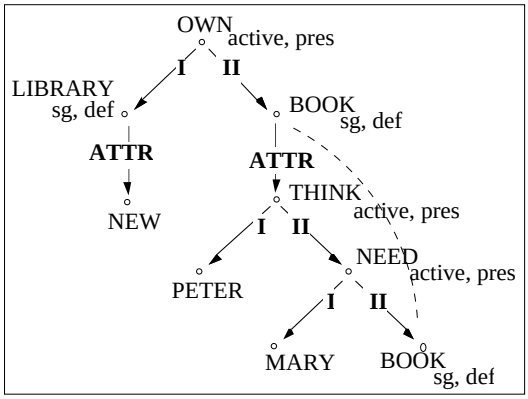

Figure 2 : DSyntS

The dictionary encodes for each generalized lexeme the associated semanteme along with the correspondence between Sem arguments and DSynt arguments.
Notation : the word *library* is a form of the lexeme LIBRARY whose semanteme is 'library'.

2. The DT nodes as semantemes

We assume the following linguistic properties for

[1] (RJ92) relate the DT to the deep syntactic structure (DSyntS) of MTT, namely a syntactic dependency tree, but they note that this correspondence DT / DSyntS is not direct, because the interpretation of adjunction arcs in terms of dependencies is not constant. (RVW95) take this divergence between DT and dependency tree as one of the motivations for defining D-Tree Grammars.

[2] We are thankful to Anne Abeillé, Laurence Danlos and Owen Rambow for valuable comments on earlier versions of this work.

elementary trees. The elementary trees correspond to exactly one semantic unit (A91)[3], and respect the predicate-argument co-coccurrence principle (PACP), though with a semantic interpretation : semantic predicates anchor trees with positions for the syntactic expression of *all and only* their semantic arguments.[4] These positions are typed as substitution nodes and foot nodes. For instance in the tree for an attributive adjective, the adjective semantically governs the semanteme represented by the foot node.[5] Traditionally auxiliary trees are used for recursive structures. If syntactic structure is considered though, another dichotomy cuts across the distinction initial/auxiliary: the syntactic head is either the main anchor (for *predicative* trees) or the foot node (for *modifier* trees) ((K89), (SS94)).[6] All initial trees are predicative. Typical predicative auxiliary trees are the trees for bridge verbs. [7]

Let us now compare DT nodes with SemS nodes. The DT refer to lexicalized elementary trees, which correspond to a semantic unit (cf supra). Therefore, a DT node can be conceived as a semanteme, plus information for a particular lexicalization of that semanteme and for a particular syntactic construction. Yet with respect to SemS nodes, two differences appear. First, in the DT, there can be several nodes in coreference (though this coreference is not handled by the TAG formalism), that would be represented by a single node in the SemS. And second, semantic units realized in the language as inflections (eg. number, tense ...) are represented as features in TAG and, thus do not appear as nodes of the DT. So provided inflectional semantemes are not taken into account and coreferent nodes in the DT are considered a single node, there is a one-to-one relation between the SemS nodes and the DT nodes.

3. The DT arcs as semantic dependencies

As we said previously, several works have noted divergences between syntactic dependencies and DT arcs. Our claim is that a constant interpretation of the DT arcs can be found, though in terms of *semantic* and not syntactic dependencies : substitution and adjunction arcs both represent semantic dependencies, though in the opposite direction (Fig. 3).[8] For illustration see Fig. 4 the DT and SemS for sentence (1).[9]

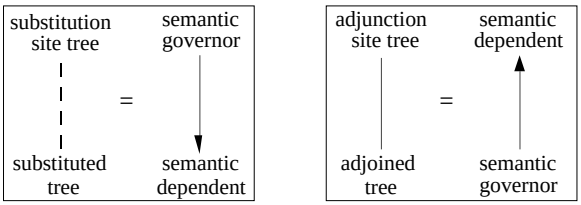

Figure 3 : Interpretation of DT arcs in terms of semantic dependencies

This result is a direct consequence of the linguistic properties we have assumed for the elementary trees. It can be noted that it is true for any type of adjunction arc (either predicative or modifier), with the definition of TAG derivation of (SS94), where multiple modifier adjunctions are allowed at the same address.[10]

---

[3] Thus elementary trees can have several lexical anchors, either because some are semantically empty (empty prepositions, complementizers ...), or because the several anchors form an idiom, whose semantic is not compositional.

[4] This counts for expressed semantic arguments only, so not for the agent in agentless passive constructions for instance.

[5] The notion of semantic governor must not be confused with the notion of *semantic head*. In « white car » white semantically governs car, yet car is the semantic head (a white car is a car). Following (P90) we define the semantic head as the semanteme that summarizes a semantic sub-graph. Not all sub-graphs can be summarized. In general a semantic graph for a whole sentence does not have a single semantic head, but one for its theme and one for its rheme.

[6] We follow the terminology of (SS94). Here *predicative* is used with its syntactic meaning.

[7] Another example is the tree for *glass of* in a *glass of wine*. The anchor *glass* is the syntactic head of the whole tree (A93). Yet the semantic interpretation of the trees for a bridge verb and for *glass-of* differ crucially: from the semantic point of view *glass of* behaves as a modifier and is not the semantic head of *glass of wine*. In *want to stay,* which expresses a will, the syntactic head is *want*.

[8] The fact that the DT should represent semantics is not new. See for example (A93) who distinguishes between *glass* in *a wine glass* and in *a glass of wine* on purely semantic grounds; (K89) who mentions that TAG should "preserve a straightforward compositional semantics"; (D98) who descirbes G-TAG, a generation system based on TAG where a derivation tree is built by lexicalizing a conceptual structure.

[9] The TAG analysis is from (X95), except that determiners are not considered as nominal complements and are thus adjoined.

[10] In case of adjunction, the interpretation in terms of semantic dependency is valid only if adjunction occurs on the spine of the tree receiving adjunction. This is the case most of the time. Yet we thank Martine Smets for pointing to us a problematic case: in *Paul gives flowers only to Mary*, *to* is semantically empty and appears as co-head in the *give* tree. The adverb *only* adjoins on the PP node of the *give* tree though it semantically governs *Mary*.

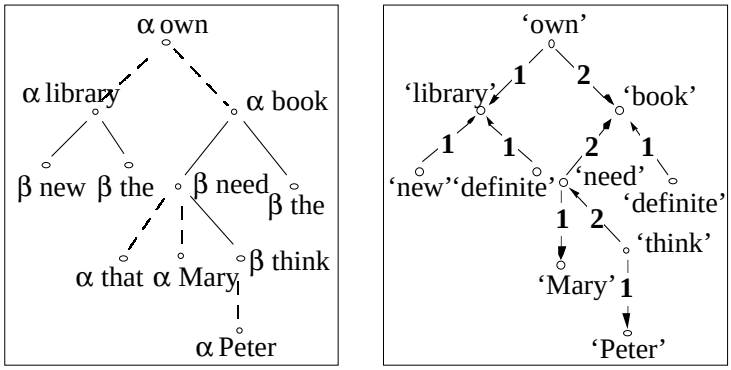

Figure 4 : DT (left) and SemS (right), with a different lay out to facilitate comparison

But obviously, the predicative adjunction arcs and the modifier adjunction arcs do not behave in the same way with respect to *syntactic* dependencies. Typically modifiers show a semantic and syntactic dependency in the opposite direction, while complement auxiliary tree preserve the direction of dependency in the semantic-syntax interface. The interaction of the various links can cause differences between the DT and the DSyntS.

Another example of mismatch is shown Fig. 4. The DT for sentence (1) shows the right chain of semantic dependencies for the sequence think-need-book, as the SemS shows. The only difference is the extra node for *that* in the DT, which does not count as a semantic unit. On the contrary in the DSyntS (Fig. 2), a syntactic dependency appears between BOOK and THINK, without a corresponding semantic dependency.

So, we have seen that in the general case, a DT induces a SemS. Further, the DT contains an additional information since it defines a partial order on its nodes, so that it form a tree. Thus the DT defines a path to cover all nodes once. The TAG procedure, from a generation point of view, is equivalent to fixing a starting node, the DT root. From that root, semantic dependencies gone through from the governor to the dependent (= positively) give substitution arcs, and semantic dependencies gone through in the opposite direction (= negatively) give adjunction arcs. It can be noted that it types the elementary trees involved as initial/auxiliary. For example, in Fig. 1, if we want to represent 'own' as a verb with two nominal arguments extended by substitution, the structure for 'think' will necessarily be an auxiliary tree, since one of its leaving arc has to be gone through negatively. Thus this gives another proof that bridge verbs have to be represented by auxiliary trees in relative clauses (or embedded interrogative clauses).

For the same reasons, to derive (4) *John knows the city in which Mary met Peter* and read the DT as a semantic graph (see the corresponding SemS Fig. 5), if the arguments of know are to be substituted, then *in* has to adjoin on *city* and *met* to substitute in *in*, though *met* is the syntactic governor of *in*.

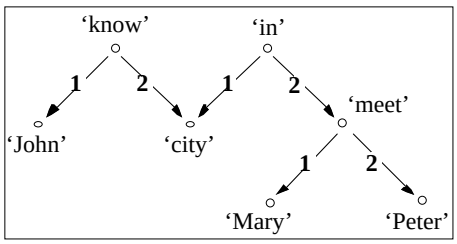

Figure 5 : SemS of (4)

4. Problematic derivations

It remains to study cases where there exists a SemS but no satisfactory DT. First TAG imposes a formal constraint that the DT be a tree. This implies in the case of cycles in the SemS, either to discard some dependency, or to cut the cycle at some node and to split that node into several coreferent ones (cf Section 2). And second, even provided a tree-like path exists for a given SemS, there are well-known cases where pure TAG fails to derive the correct word order (eg. clitic climbing in Romance (B98), or Kashmiri wh-extraction (cf RVW95)). To get the right word order a less restrictive formalism must be used.

More problematic are cases of TAG derivations showing the wrong dependencies. While adjunction of bridge verbs gives the right semantic dependencies in case of extraction, these adjunctions may be problematic when the bridge verb serves as argument for another predicate. Consider the following sentences, where a clause containing an embedded clause serves as argument for the main verb:

(5a) *Paul claims Mary said Peter left.*

(5b) *Paul claims Mary seems to adore hotdogs* (RVW95)

(5c) *That Paul wanted to stay surprised Mary.*

For (5a), in the classic TAG analysis (X95), the two bridge verbs adjoin recursively, and the DT is perfect (with the interpretation of adjunction arcs defined in Fig. 3). Yet for (5b) *Mary seems to adore hotdogs* serves as argument for *claims*, but here *seems* adjoins

on VP, and thus *claims* has to adjoin on *adore*.[11] Thus the DT does not show the right dependencies (either semantic or syntactic, cf (RVW95)). For (5c), the verb *surprised* traditionnally receives its subject via substitution (to block extraction), thus if the bridge verb *wanted* is still adjoined, the DT is different from the SemS (Fig. 6) (apart from the splitting of the 'Paul' node into 2 coreferent nodes; we show the coreference with a curved dashed line). The problem arises because the tree αstay substitutes in αsurprise, but when the predicative tree βwant adjoins on αstay, it becomes the semantic head of the whole subtree.[12]

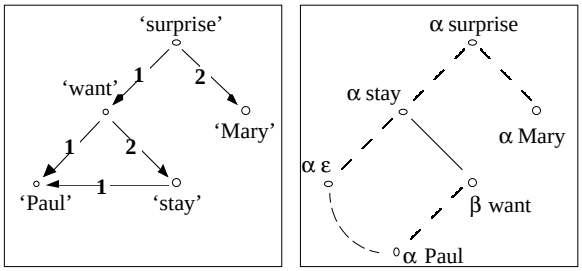

Figure 6 : Problematic derivation (SemS and DT) for
That Paul wanted to stay surprised Mary

So to read a DT as a SemS, we need not only the PACP, but also a control over the combination of the elementary trees : it must be checked that the argumental positions in a tree are actually filled by the right arguments.[13]
It can be noted that for sentence (5b) and (5c), ruling out adjunctions of complement trees (as in DTG (RVW95)) solves the problem. Yet it might be problematic for sentence (1), for which we have seen that the TAG DT shows the right semantic dependencies. And it also rules out the adjunction of an athematic complement tree (such as the one for *glass-of*). This is investigated in (CK98).

Conclusion

We have shown that in the general case the DT can be viewed as a semantic representation, in the sense of MTT, provided coreference is not taken into account. We have given a characterization of problematic derivations. This result is of crucial importance for any further processing based on the TAG derivation tree.
We have also provided a new characterization of adjunction and substitution arcs depending on the direction of the semantic dependency they represent.

[11] (SS94) already noted that multiple adjunctions of bridge verbs at one node should be ruled out, here we find that this holds for a whole tree.

[12] (K89) already noted that « derivations under which thematic roles, once established, are altered by further adjunctions » should be ruled out.

[13] Another case where positions « are not filled by the right arguments » is for instance pied-piping. The XTAG derivation for *the woman whose daughter Peter talks to* does not show the right semantic dependencies, since a link appears between *talks-to* and *woman*.

References

(A91) A. Abeillé, 1991: *Une grammaire lexicalisée d'arbres adjoints pour le français*. Ph.D. Thesis. Univ. Paris 7.

(A93) A. Abeillé, 1993: Interactions Syntaxe-Sémantique dans une TAG. Colloque ILN'93, Nantes.

(B98) T. Bleam, forthcoming: Clitic Climbing and the power of TAG, in Abeillé, A., Rambow, O. (eds.), *Tree-adjoining Grammars*, CSLI, Stanford.

(CK98) M-H. Candito, S. Kahane, 1998: Defining DTG derivations to get semantic graphs. This volume.

(D98) L. Danlos, forthcoming, G-TAG : a formalism for text generation, in Abeillé, A., Rambow, O. (eds.), *Tree-adjoining Grammars*, CSLI, Stanford.

(F92) R. Frank, 1992: *Syntactic locality and Tree Adjoining Grammar: Grammatical, Acquisition and Processing Perpectives*. Ph.D. Thesis. Univ. of Pennsylvania.

(K89) A. Kroch, 1989: Asymmetries in Long-Distance Extraction in a Tree-Adjoining Grammar. In *Alternative Conceptions of phrase structure*. M. Baltin and A. Kroch (eds), Univ. of Chicago Press, Chicago.

(M88) I. Mel'cuk, 1988: *Dependency Syntax : Theory and Practice*. Albany. State Univ. of NY Press.

(P90) A. Polguère, 1990 : *Structuration et mise en jeu procédurale d'un modèle linguistique déclaratif dans un cadre de génération de texte*. Ph.D. Thesis. Univ. of Montreal.

(RJ92) O. Rambow, A. Joshi, 1992: A Formal Look at Dependency Grammars and Phrase-Structure Grammars, with Special Consideration of Word-Order Phenomena. in Leo Wanner (ed.) *Recent Trends in Meaning-Text Theory*

(T59) L. Tesnière, 1959: *Eléments de syntaxe structurale*, Klincksieck, Paris.

(RVW95) O. Rambow, K. Vijay-Shanker, D. Weir, 1995: D-Tree Grammars, ACL'95.

(SS94) S. Shieber, Y. Schabes, 1994: An alternative conception of tree-adjoining derivation. *Computational Linguistics*, 20.1.

(XTAG95) XTAG research group: A Lexicalized TAG for English. Technical report IRCS 95-03, Univ. of Pennsylvania. (Online updated version).

(ZM67) A. Zolkovskij, I. Mel'cuk, 1967 : O semanticeskom sinteze [On semantic synthesis]. *Problemy kybernetiki*, v. 19, 177-238.