Wh-dependencies in Romanian and TAG
Manuela Leahu
TALANA, Université Paris 2, place Jussieu, case 7003, 75251 Paris cedex 05
manuel@pirae.linguist.jussieu.fr

Introduction

The aim of this paper is to analyse the wh-movement for Romanian in TAG formalism. Romanian shares free extractibility from tensed clauses with its Romance sister languages and it has borrowed multiple wh-fronting from the slavic languages. These features of Romanian are quoted by Kroch (1989) from Comorovsky (1986), where he justifies the analysis of extractions in TAG. This formalism allows a correlation between the absence of wh-islands and the possibility of multiple wh-movement. But the facts of the Romanian language are more complex. We consider here several phenomena like simple questions, unbounded dependencies, wh-islands, multiple wh-movement. Because of order between the free wh-words for the multiple wh-movement, a complete analysis is not possible with TAG.

TAG derivation trees do not provide a good representation of the dependencies between the words of the sentence, i.e., of the predicate–argument and modification structure. Also, the derivation structures of MCTAG (Joshi, 1987) cannot be given a linguistically meaningful interpretation (Section 3). We show here that an analysis is possible with DTG formalism (Vijay-Shanker, D. Weir, O. Rambow, 1995) that resolve these problems with the use of a single operation—that we call subinsertion—for handling all complementation.¹

Simple Questions

(1) Pe cine vede Ion e?
Who sees Ion e?

This sentence in the TAG formalism is represented as a transitive tree with object extraction and the initial place of the extraction is marked by a trace. A characteristic feature for questions is the inversion of the subject.

Unbounded dependencies

The following sentences illustrate some examples of unbounded dependencies:

(2) Ce regreti ca a citit Maria e?
What do you regret that Mary has read e?

The wh-pronoun in the initial tree is in the same verb with which it is construed and its interpretation as the object of the verb “to read” is thus guaranteed. Following standard conventions we represent the relationship between the fronted constituent and the position in which phrases with its grammatical role normally appear by coindexing the fronted wh with an empty category. The relationship between an indexed empty category and the categorially identical, c-commanding node with which it is in coindexed, we call “linking.” The adjunction of the auxiliary tree in the initial tree produces the final tree in which the wh-word is now initial in the matrix sentence. Strikingly, there is no bound on the depth of the embedding:

(3) Pe cine crezi ca Paul a zis ca Ion a placut e?
Who do you think that Paul said that Ion liked e?

In (3), the wh-word is an argument of the most deeply embedded verb “like”, thus causing the non-projectivity. A TAG can capture the long-distance dependency naturally, since the recursive adjunction operation allows an unbounded number of clauses to intervene between directly dependent lexemes. We first substitute all nominal arguments into their respective verbal trees, and then adjoin the intermediate say-clause into the most deeply embedded like-clause at the S node immediately dominated by the root. This has the effect of separating the wh-word from its verb, even though they originated in the same structure. We than subsequently adjoin the matrix think-clause into the intermediate say-clause.

¹ We are grateful to Anne Abeille, Sylvain Kahane and Owen Rambow.
Wh-islands

Islands phenomena can be found in Romanian for relative clauses and adjuncts.

(4) *Pe cine j cunosti femeia i care a intalnit e1?
Who do you know the woman which met e1?

(5) *Pe cine j ai plecat inainte sa examineze e1 Ionescu?
Who did you leave before that examine e1 Ionescu?

These violate for locality reasons: there is no way to localize the wh-element and its co-indexed base position in the same tree set (MCTAGs) which can then be adjoined into a single elementary tree.

But in the case of interrogative clauses are not islands for extraction:

(6) Pe cine j crezi ca Paul detesta e1?
Who do you think that Paul detests e1?

(7) Pentru care clauza vrei sa afli e1;e; nu a decis inca ce va vota e1 e2? (Comorovsky 1986)
For which paragraph do you want to learn whoiei has not decided yet what he will vote e1 e2?

For this Kroch suggests an interesting account that reduces the constraint on movement out of an island to a local well-formedness condition on elementary trees.

Multiple Wh-Movement

In Romanian language multiple wh movement are rare but grammatical. Wh pronouns are strictly ordered:

(8a) Cine cui j promite o masina e1?
To whom whom promises a car e1?

(8b) *Cui j cine promite o masina e1?
To whom who promises a car e1?

(9a) Cui j ce j zice e1 ca vezi e1?
To whom what he says e1 you see e1?

(9b) *Ce j cui j zice e1 ca vezi e1?

(10a) Cui j pe cine j Paul zice e1 ca vezi e1?
To whom whom Paul says e1 you see e1?

(10b) *Pe cine j cui j Paul zice e1 ca vezi e1?

(11a) Cine j pe cine j a zis e1 ca a vazut e1?
Who whom said e1 he has seen e1?

(11b) *Pe cine j cine j a zis e1 ca a vazut e1?

(12a) Cine ce j ziceai ca e1 isi inchipuie ca ai descoperit e1?
Who, what you were saying that e1, to himself imagines that you have discovered e1?

(12b) *Ce j cine j ziceai ca e1 isi inchipuie ca ai descoperit e1?

(13) Cine j cui j ce j ziceai ca e1 i-a promis e1 e1 j?
Who to whom what you were saying that e1 to him has promised e1 e1 j?

Examples 8(b)-12(b) are not correct because they don't respect the ordering on the wh-pronouns, which is the following:

cine(who)<cui (to whom)<pe cine(whom)<ce(what)

Nom<Dat<Acc

The ordering constraint is kept even if the Wh extractions are not dependent on the same verb (9-12(a)).

When a non pronominal NP is also extracted, several word orders are possible:

(14a) Ce masina i cui j Paul promite e1 sa repare e1?
What car to whom Paul promises e1 to repair e1?

(14b) Cui ce masina j Paul promite e1 sa repare e1?

To derive the sentence in (14a), for example, we adjoin the tree "to whom Paul promises" into the elementary tree "which car to repair" and this example can be analysed in TAG formalism. (Figure 1)

![Figure 1: trees for (14a)](image-url)

In examples 9-12(a) and 14(b) because of the ordering constraint, the TAG formalism is not able to analyse these cases, given the predicate-argument cooccurrence constraint on elemen-
tary trees.
The problem in describing this phenomena with TAG arises from the fact observed by Vijay-Shanker 1992, that adjoining is an overly restricted way of combining structures.

In Multi-Component TAG (MCTAG) (Joshi, 1987), trees are grouped into sets which must be adjoined together (multicomponent adjunction). The elementary tree is split up into parts, which are grouped together into sets. All trees from one set must be adjoined at the same time, at different nodes into the single tree representing the embedded clause. However, MCTAG lack expressive power since, while syntactic relations are invariably subject to c-command or dominance constraints, there is no way to state that two trees from a set must be in a dominance relation in the derived tree. (Figure 2)

DTG is designed to overcome this limitation. Subsertion can be viewed as a generalization of adjunction in which components of the clausal complement (the subserted structure) can be interspersed within the structure that is the site of the subsertion. DTG provide a mechanism involving the use of domination links (d-edges) that ensure that parts of the subserted structure that are not substituted dominate those parts that are. Furthermore, there is a need to constrain the way in which the non-substituted components can be interspersed.

The derivation proceeds as follows: we first subsert the embedded clause tree into the matrix clause tree. After we subsert the wh-pronoun of the first clause and the wh-pronoun of the second clause (The extraction of the first clause precedes the extraction of the second clause) (Figure 3)

Figure 2: MC-trees for (14b)

In DTG formalism, the ordering constraint on the extractions is marked by the feature “topic”. The final tree is the desired, semantically motivated, dependency structure:

![Figure 3: quasi-trees for (14b)](image-url)
the embedded clause depends on the matrix clause), with respect to the ordering constraint on the wh-pronouns.

**Conclusion**

DTG are designed to share some of the advantages like other formalisms in the TAG family, while overcoming some of their limitations. The most distinctive feature of DTG is that there is complete uniformity in the way that the subinsertion relate lexical items. Furthermore, DTG can provide a uniform analysis for wh-movement in Romanian, despite the fact that the wh-elements in Romanian can appear in sentence-initial position and in sentence-second position.

**References:**

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

A. Abeillé, forthcoming: Extraction out of NP and clitic-noun dependencies in French. TAG+3, Paris.

I. Comorovsky, 1986: Multiple WH Movement in Romanian, Linguistic Inquiry.

C. Dobrovie, 1994: The syntax of Romanian. Mouton de Gruyter.

R. Frank, 1992: Syntactic Locality and TAG: Grammatical, Acquisition and Processing Perspectives. Ph.D. thesis. Univ. of Pennsylvania.

A. Joshi, O. Rambow, 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.

A. Kroch 1989: Asymmetries in Long Distance Extraction in a TAG. In Alternative conceptions of phrase structure. M. Baltin and A. Kroch (eds), Univ. of Chicago Press, Chicago.

A. Kroch and A. Joshi 1986: Analysing extraposition in a TAG. In G. Huck and A. Ojeda (eds), Syntax and Semantics: Discontinuous Constituents.

O. Rambow, K. Vijay-Shanker, D. Weir. 1995: D-Tree Grammars, ACL 1995.

XTAG for English 1995, Institute for Research in Cognitive Science, Univ. of Pennsylvania.