

# Cross-Serial Dependencies in Tagalog

Anna Maclachlan and Owen Rambow

*Discern Communications and University of Pennsylvania*

anna@discerncomm.com and rambow@unagi.cis.upenn.edu

## 1. Cross-Serial Dependencies

There are two salient linguistic uses of adjunction: for analyzing long-distance *wh*-movement (and related movement types) in many languages and for analyzing cross-serial dependencies (CSD) in Dutch and Swiss German. While the need for and the adequacy of adjunction to model *wh*-type movement have been questioned (Rambow and Vijay-Shanker, 1998; Rambow et al., 2001), CSD seems ideally suited for a TAG analysis, since, as Shieber (1985) showed, CSD cannot be derived by a context-free grammar. In fact, some of the alternate tree rewriting systems proposed which do not include adjunction, such as the DSG of (Rambow et al., 2001), cannot provide a satisfactory analysis of CSD, either: it is specifically the definition of adjunction as an tree-rewriting operation that inserts one tree in its entirety into the center of another that is crucial for deriving CSD. What is somewhat troubling, however, is that the construction appears to be limited to two West Germanic languages/dialects, Dutch and Swiss German. In this paper, we show that the same construction, though with different syntactic characteristics, is found in a completely unrelated language, Tagalog. We show how the analysis of Kroch and Santorini (1991) for Dutch can be adapted for Tagalog, and we show furthermore that the syntactic analysis suggested by TAG is preferable to an analysis based on head movement and verb incorporation.

## 2. The Tagalog Data

Tagalog, a major Austronesian language spoken in the Philippines, is strongly verb first. Complements and the subject follow the verb with preferences for the agent to directly follow the verb and for the nominative argument to be last (preferences which can be in conflict). The nominals are case marked for nominative (NOM) and oblique (OBL), and another distinguished case is un-glossed in the examples (for a discussion of this case as both ergative and accusative see (Maclachlan, 1994)). The standard ordering in complex sentences is V1 Agent1 linker [V2 (Agent2) Theme2], as shown in example (1a). Phrases of various sorts are separated by a linker (LK) and Tagalog also has sentential conjunction (CONJ). A cross serial dependency ordering alternates with this basic ordering in which the agent of the matrix clause follows the embedded verb as in (1b): V1 linker V2 Agent1 (Agent2) Theme2.<sup>1</sup>

### (1) Basic and CSD alternates

a. Nagisip si Pedro-ng bumili ng bulaklak  
AT-thought NOM-Pedro-LK AT-buy flower

b. Nagisip na bumili si Pedro ng bulaklak  
AT-thought LK AT-buy NOM-Pedro flower

‘Pedro thought to buy (of buying) a flower.’

Let us note two further properties of the CSD for which we will account with a TAG analysis. First, the CSD process can be iterated:

### (2) Iteration of CSD

---

1. Baldridge (1998) claims that Tagalog simply has long-distance scrambling and that a  $V_1 V_2 N_2 N_1$  ordering is also allowable. However, he gives only one example, and admits that the sentence may have a completely different interpretation (in which the two NPs form one NP). We have, in our work with native speaker informants, not found any evidence for generalized long-distance scrambling, and therefore will assume for the sake of this paper that we have a CSD, not a long-distance scrambling construction. If it were in fact a long-distance scrambling construction, TAG would not be powerful enough for an analysis – see (Rambow, 1994) for a discussion of German. Baldridge (1998) also discusses asymmetries in *wh*-extraction in Tagalog. We do not address the issue here.

- a. Ipiniangako ni Maria-ng subuka-ng manalo sa karera  
promised Maria-LK try-LK win OBL-race
- b. Ipiniangako-ng subukan ni Maria-ng manalo sa karera  
promised-LK try Maria-LK win OBL-race
- c. Ipiniangako-ng subuka-ng manalo ni Maria sa karera  
promised-LK try-LK win Maria OBL-race  
'Maria promised to try to win the race.'

Second, the CSD sentence permits only one NOM nominal, while the basic complex sentence permits two. This can be seen when the theme is NOM in the embedded clause as in (3). In this essentially passive clause type, the verb is marked with Theme Topic morphology (TT) whereas in the essentially active clause type, as in both clauses in (1), the verb is also marked but with Agent Topic morphology (AT). While the matrix agent is NOM in (3a) it cannot be in the CSD equivalent in (3b) as long as the embedded theme is NOM

(3) Basic and CSD alternates with embedded passivization

- a. umasa si Maria-ng sulatin ang kuwento  
AT-hoped NOM-Maria-LK write-TT NOM-story
- b. umasa-ng sulatin {\*si Maria / ni Maria} ang kuwento  
AT-hoped-LK write-TT {\*NOM-Maria / Maria} NOM-story  
'Maria hoped to write the story.'

### 3. A TAG Analysis

In TAG, we derive CSD by recursively adjoining elementary trees into each other at interior nodes. As is the case with all embedded clause constructions (be they CSD or not), each clause is adjoined into its immediately embedded clause, since the most deeply embedded clause does not have a linguistically meaningful footnote labeled with a clausal category (and hence its embedding clause must adjoin into it, rather than *vice versa*). When we adjoin an auxiliary elementary tree such as that shown schematically at the top left in Figure 1 (the superscript 1 indicates that this represents the matrix clause) to the initial tree at the top-right (the superscript 2 indicates that this is the embedded clause, we are only considering one level of embedding in this schematic discussion, though of course the process can iterate), the result is as shown below in Figure 1. The nodes labeled *A*, *B*, *C*, *D* represent either substitution or terminal nodes – in either case, these are positions below which terminals can be generated. Of course, we do not expect terminal symbols to be generated below each of these symbols. In fact, if we restrict ourselves to the case in which we have one (overt) noun phrase and one verb in each elementary tree, two of the symbols will dominate the empty string. If we choose *A* and *D* to dominate the empty string, we obtain a center-embedded structure with the associated string  $B^{(1)}B^{(2)}C^{(2)}C^{(1)}$ , as desired. The derivation is essentially a context-free derivation and does not actually make use of the full power of adjunction, since no terminal nodes are generated above the adjunction site.

It is clear that to obtain CSD, we must choose as overt terminal nodes one above the adjunction site (*A* or *D*) and one below (*B* or *C*). If we choose, say, *A* and *B* as the overt nodes, we obtain a structure which is not derivable with a context-free grammar, but the string nonetheless represents center-embedding ( $A^{(2)}A^{(1)}B^{(1)}B^{(2)}$ ). Thus, we must choose one overt terminal to the left of the spine, and one to the right. This leaves us with exactly two possibilities – *A* and *C* are overt, or *B* and *D*. Since in Dutch and Swiss German CSD, the first element is always a matrix noun phrase (and not an embedded one), we cannot use *A* and *C* as the overt elements: while adjoining the matrix clause into the embedded clause would result in  $A^{(2)}A^{(1)}C^{(2)}C^{(1)}$ , with cross-serial dependencies, the string starts with the wrong *A*:  $A^{(2)}$  rather than  $A^{(1)}$ . Thus, we must leave *A* and *C* empty, with the overt material in *B* and *D*.

This is of course exactly the choice that Kroch and Santorini (1991) make. They propose that in Dutch,<sup>2</sup> the verb raises from its ordinary position as sister to the *S* footnote to a position above the adjunction site (which

2. Their analysis also applies to the relevant Swiss German data.

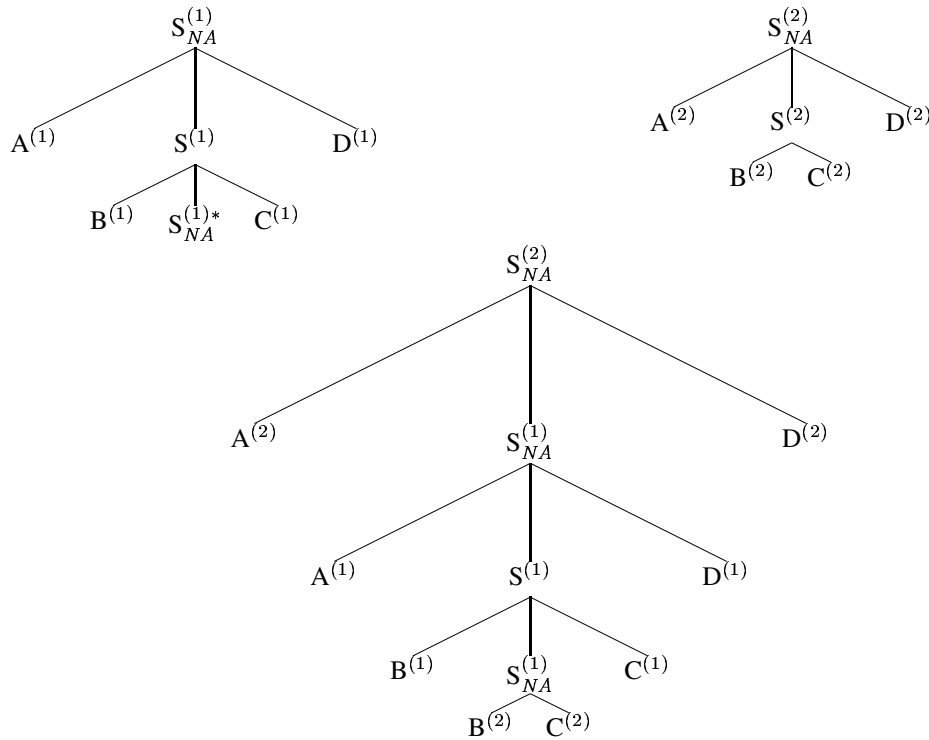


Figure 1: Elementary trees (auxiliary, above left, and initial, above right) and derived tree obtained by adjoining the auxiliary tree into the initial tree at the latter's interior  $S$  node (below)

can be interpreted as right-Chomsky-adjoining<sup>3</sup> to the regular maximal projection). What makes their analysis so compelling is that this analysis, in which the verb “raises” to a higher position in the tree, takes up some elements of the analysis suggested previously in the Germanic syntax literature. In this analysis, which dates to at least Bech (1955) and was expressed somewhat more formally in a transformational framework by Evers (1975), the verbs actually raise out of their clauses and form a single morphological unit. Such an analysis is impossible in TAG, since apart from the effect of adjunction, the elementary trees retain their structural integrity. Furthermore, Kroch and Santorini (1991) argue that there are empirical arguments against a morphological verb cluster, though not against verb raising itself. Thus, the analysis proposed by Kroch and Santorini (1991) is the closest possible TAG-based analysis which uses the independently proposed notion of verb raising (but not verb cluster formation), and it also corresponds to the only possible analysis considering the topology of trees and the definition of adjunction!

Let us now turn to Tagalog. In Tagalog, we have a verb-initial construction rather than a verb-final construction. However, the argument about possible analyses is exactly the same as in Dutch, and we conclude that  $B$  and  $D$  must be overt, not  $A$  and  $C$ . Because Tagalog is verb-initial, we must choose  $B$  to represent the verb, and  $D$  to represent the noun phrase. We thus are forced to adopt an analysis in which the NP is raised, and in which it is the raising of the NP which results in the CSD.<sup>4</sup> This is shown in Figure 2 (the subscripts indicate the relation between traces and moved elements within elementary trees, while the superscripts, as before, indicate which clause a terminal symbol belongs to). The trees in this figure derive the CSD version of (1), (1b), repeated here for convenience:

(4) Basic and CSD alternates (=1))

- a. Nagisip si Pedro-ng bumili ng bulaklak  
AT-thought NOM-Pedro-LK AT-buy flower

3. we use “Chomsky-adjoining” to refer to derivation processes within elementary trees (following the general approach of (Frank, 2001)), while “adjoining” refers to the TAG operation that combines elementary trees.

4. This does not mean that the verb cannot also raise from a VP-internal position to a higher position on its own projection, as is customarily assumed for verb-initial languages. It just means that the landing site of the verb must be below the node at which adjunction of the matrix clause happens. This is in fact exactly the analysis in Figure 2.

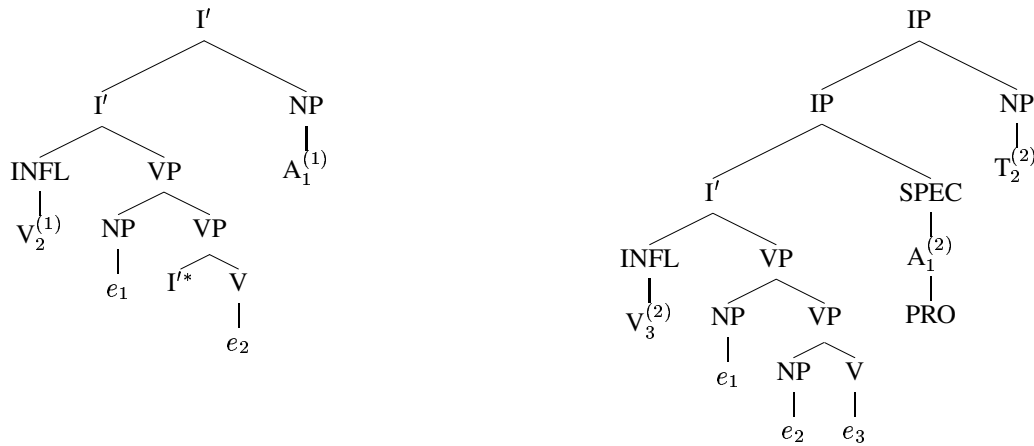


Figure 2: Elementary trees for Tagalog clauses in CSD order: matrix clause (left) and embedded clause (right)

- b. Nagisip na bumili si Pedro ng bulaklak  
 AT-thought LK AT-buy NOM-Pedro flower

‘Pedro thought to buy (of buying) a flower.’

We assume that Tagalog is underlyingly verb-final,<sup>5</sup> and in both trees the verb has raised from V to INFL. In the matrix clause on the left (with superscript (1)), the verb also subcategorizes for an I' (the footnode) and an NP, the agent, which has raised to a position to the right of the spine, Chomsky-adjoining to I'. In the embedded clause on the right, the verb subcategorizes for two NPs, the first of which (the agent) is realized as PRO and occupies the SPEC of IP position, and the second of which (the theme) has scrambled out beyond the agent. The matrix clause adjoins to the matrix clause at its I' node, and the result is the derived tree shown in Figure 3, with the word order in sentence (4b). We will postpone a discussion of the non-CSD case (4a) until after a discussion of case assignment.

The question arises whether this analysis, imposed on us as it is by the definition of adjunction, is independently motivated. As it turns out, in Tagalog there is independent linguistic evidence for NP raising (just as there is independent linguistic evidence for verb raising in Dutch). The evidence comes from the agreement facts in Tagalog. In the following, we will assume that NOM (a case marker which does not fully correspond to nominative case in other languages) is assigned through SPEC-head agreement, and that IP is SPEC-final.<sup>6</sup> Rather than being associated with tense or aspect marking, NOM is associated with topic morphology (TT and AT) in Tagalog. We will assume that other, non-NOM arguments (including the clausal argument) may optionally leave the VP and Chomsky-adjoin to the I' or IP nodes (without being assigned NOM). This optional movement of arguments is at the heart of the availability of two analyses in Tagalog: we obtain the “basic” word order when the clausal argument has been moved beyond the NOM argument, as shown in Figure 4 on the left. We will assume that when the clausal argument has not moved out of the VP, it cannot project to more than I', presumably for reasons related to case (or some extended notion of case). This will prevent us from deriving nested dependencies. At the same time, this asymmetry – between a node label of IP for the basic word order, and a node label of I' for the CSD word order – reflect a widely held intuition that in the CSD order, the clauses are less “separated” from one another than in the basic word order. Similar intuitions in Dutch (and other West Germanic languages) led Bech (1955) and Evers (1975) to postulate the existence of only one clause in certain constructions (“clause union”).

We now turn to the question why in the CSD construction, there can be at most one NOM-marked nominal argument, while in the basic construction, each clause may have its own. It is clear that in our analysis of the basic construction, there are two IPs each with its own SPEC position, so that two NOM cases can be assigned independently, as in (3a), repeated here for convenience.

- (5) Basic and CSD alternates with embedded passivization (=3)

5. This assumption is actually irrelevant from the TAG point of view, but it is consistent with much recent work on the syntax of V1 languages.

6. There is cross-linguistic support for this claim from other Austronesian languages such as Malagasy.

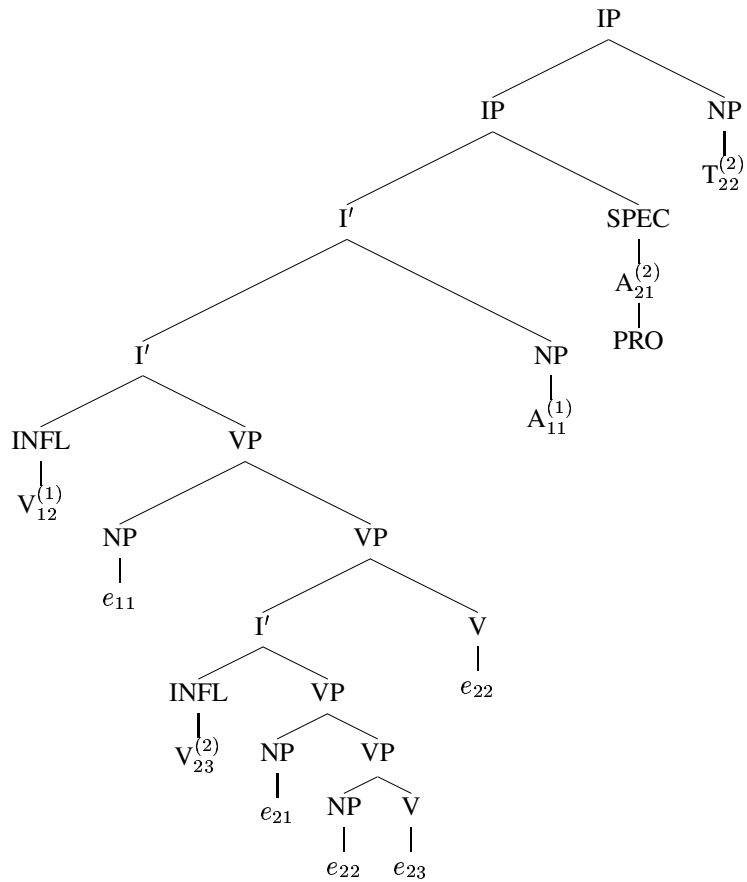


Figure 3: Derived tree for Tagalog clauses in CSD order

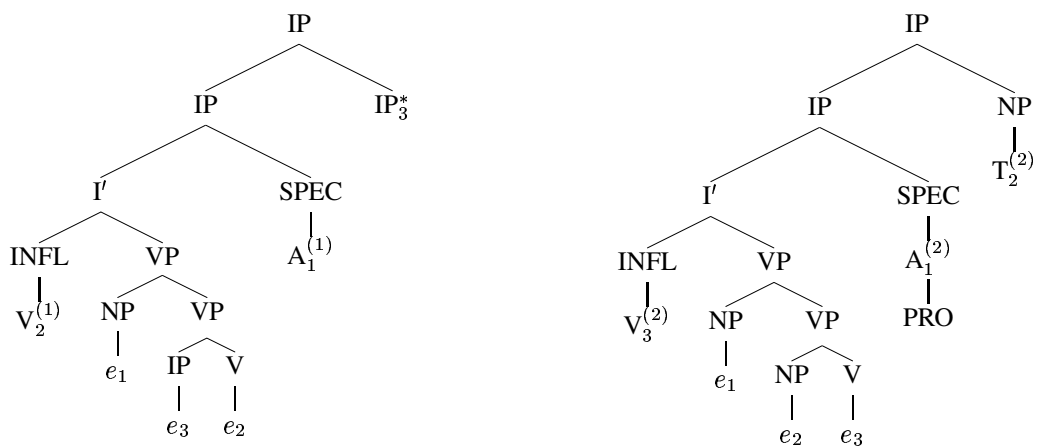


Figure 4: Elementary trees for Tagalog clauses in standard order: matrix clause (left) and embedded clause (right)

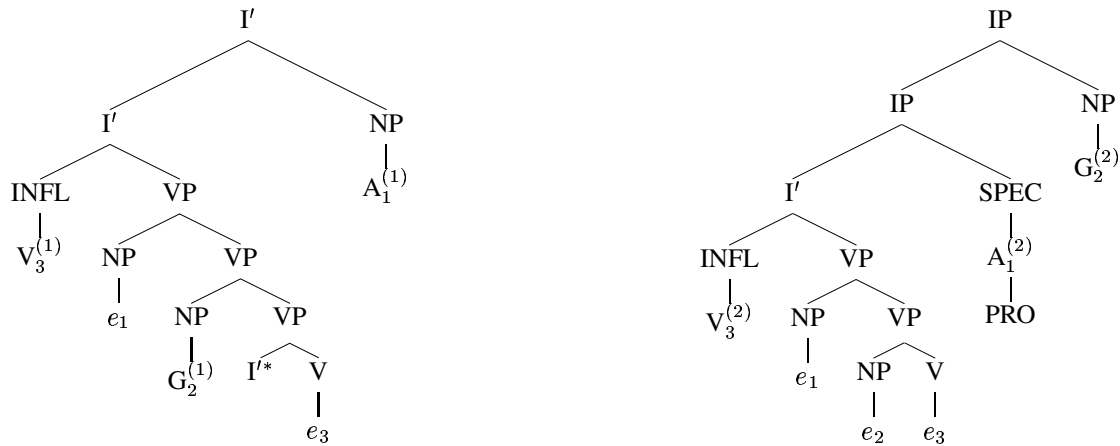


Figure 5: Elementary trees for Tagalog clauses in CSD order with interleaved nominal arguments, corresponding to sentence (7): matrix clause (left) and embedded clause (right)

- a. *umasa si Maria-ng sulatin ang kuwento*  
 AT-hoped NOM-Maria-LK write-TT NOM-story
- b. *umasa-ng sulatin {\*si Maria / ni Maria} ang kuwento*  
 AT-hoped-LK write-TT {\*NOM-Maria / Maria} NOM-story
- ‘Maria hoped to write the story.’

We might be tempted to derive the fact that the matrix clause agent cannot receive NOM case along with the embedded verb theme (as shown in (5b)) from the fact that the matrix clause simply does not have a SPEC position for INFL, and therefore it cannot have a NOM-marked argument. However, this cannot be the whole story, since it is clearly wrong: in (4) we have a NOM-marked matrix argument in both the basic and CSD word orders. A different approach would be to postulate the existence of a feature shared across and between projections of different verbs, but not across IP nodes (again capturing the same intuition of “clause union”). This feature would ensure that only one constituent is marked NOM. However, this theory misses an additional complicating factor: in (5), it is altogether impossible for the matrix agent to be NOM-marked, whether or not the embedded theme is NOM-marked:

- (6) \* *umasa-ng sulatin si Maria {ang kuwento / ng kuwento}*  
 AT-hoped-LK write-TT NOM-Maria {NOM-story / story}

Intended meaning: ‘Maria hoped to write the story.’

We therefore return to an analysis which exploits the node labels, but we make an additional assumption: case assignment can be shared between an overt argument and a PRO it controls (just as other features, such as referential indices, are shared). Thus, the matrix agent in (4b) gets its NOM not from the matrix SPEC of IP (since there is none), but rather through co-indexation with the embedded agent, which is PRO in SPEC of IP of the embedded verb. In fact, in all cases in which we have a NOM case in a matrix clause of a CSD, it is the controller of a PRO in the lower clause which must be in SPEC of IP, given the verbal morphology. While the notion of case being passed from PRO to its controller may seem at first strikingly odd, we note that in the derived tree (see Figure 3), the PRO actually c-commands its controller!

#### 4. Evidence Against Verb Incorporation

We point out a series of cases that are readily handled in the TAG analysis which pose a problem for two other types of analyses of CSDs. As argued in (Maclachlan, 1991), analyses that assume a mechanism like head-to-head movement of an embedded verb head into the matrix verb head similar to a causative verb incorporation approach or a morphological verb complex (Evers, 1975) cannot account for these cases. It is possible for a phrase from

the matrix clause to occur intermingled with embedded clause elements even while other matrix clause elements remain in place. Namely, negation, floated quantifiers and full phrases can intervene between the two verbs of a CSD. One such example is given in (7) where the matrix clause has an agent, a sentential and an oblique argument. The basic order is V1 Agent1 Goal1 [CONJ V2 Goal2] as in (7a) but a possible CSD order is V1 *Goal1* CONJ V2 Agent1 Goal2 as in (7b) where the matrix oblique phrase remains between the verbs.

(7) Basic and CSD alternates with intervening phrase

- a. Sinabi ni Fe kay Juan kung kailan tatawag kay Maria  
TT-said Fe OBL-Juan CONJ when AT-call OBL-Maria
- b. Sinabi kay Juan kung kailan tatawag ni Fe kay Maria  
TT-said OBL-Juan CONJ when AT-call Fe OBL-Maria  
'Fe told Juan when to call Maria.'

We can derive such cases by assuming that in the analysis in (2), the matrix goal has remained in the VP, while the matrix agent is Chomsky-adjoined to  $I'$ . (Note that no NOM is assigned in this example.) This is shown in Figure 5. Crucial to our analysis is the fact that each verb has its own projection which enters into the derivation fully formed, thus allowing the intervening material.

## 5. Conclusion

In this paper, we have discussed cross-serial dependencies in Tagalog. As in the case of Dutch discussed by Kroch and Santorini (1991), the definition of adjunction leaves only one possible way of using TAG in the linguistic analysis. As in Dutch, but for completely different reasons, this analysis is independently motivated by other (non-TAG) linguistic analyses. Furthermore, as in the case of Dutch, analyses have been proposed for Tagalog which include processes such as clause union or verb complex formation/verb incorporation which cannot readily be modeled by TAG. Again, as in the case of Dutch, there is independent empirical evidence against such a process. The striking parallel between Dutch and Tagalog, despite the stark differences in syntax between the two languages, lends further credence to the claim that adjunction represents a linguistically meaningful operation.

## References

- Baldrige, Jason Michael (1998). Local scrambling and syntactic asymmetries in Tagalog. Master's thesis, University of Pennsylvania.
- Bech, Gunnar (1955). *Studien über das deutsche Verbum infinitum*. Det Kongelige Danske videnskabernes selskab. Historisk-Filosofiske Meddelelser, bd. 35, nr.2 (1955) and bd. 36, nr.6 (1957). Munksgaard, Kopenhagen. 2nd unrevised edition published 1983 by Max Niemeyer Verlag, Tübingen (Linguistische Arbeiten 139).
- Evers, Arnold (1975). *The Transformational Cycle in Dutch and German*. PhD thesis, University of Utrecht. Distributed by the Indiana University Linguistics Club.
- Frank, Robert (2001). *Phrase Structure Composition and Syntactic Dependencies*. MIT Press, Cambridge, Mass.
- Kroch, Anthony and Santorini, Beatrice (1991). The derived constituent structure of the West Germanic Verb Raising construction. In Freidin, R., editor, *Principles and parameters in comparative grammar*, pages 269–338. MIT Press, Cambridge, Mass.
- Maclachlan, Anna (1991). Implications of a Tagalog construction for incorporation theory constraints. In Wilson, Tom, editor, *Proceedings of the Annual Conference of the Canadian Linguistic Association*, pages 209–220, Toronto.
- Maclachlan, Anna (1994). Conjunction reduction and the syntax of case in Tagalog. In Ode, Cecilia and Stokhof, Wim, editors, *Proceedings of the Seventh International Conference on Austronesian Linguistics*, pages 443–460, Leiden, the Netherlands.
- Rambow, Owen (1994). *Formal and Computational Aspects of Natural Language Syntax*. PhD thesis, Department of Computer and Information Science, University of Pennsylvania, Philadelphia. Available as Technical Report 94-08 from the Institute for Research in Cognitive Science (IRCS) and also at <ftp://ftp.cis.upenn.edu/pub/rambow/thesis.ps.Z>.
- Rambow, Owen and Vijay-Shanker, K. (1998). *Wh-islands in TAG and related formalisms*. In *Proceedings of the Fourth International Workshop on Tree Adjoining Grammars and Related Frameworks (TAG+4)*, number 98–12 in IRCS Report, pages 147–150. Institute for Research in Cognitive Science, University of Pennsylvania.
- Rambow, Owen; Vijay-Shanker, K.; and Weir, David (2001). D-Tree Substitution Grammars. *Computational Linguistics*, 27(1).
- Shieber, Stuart B. (1985). Evidence against the context-freeness of natural language. *Linguistics and Philosophy*, 8:333–343.