

A Tree Adjoining Grammar Analysis of the Syntax and Semantics of *It*-Clefts

Chung-hye Han

Department of Linguistics
Simon Fraser University
chunghye@sfu.ca

Nancy Hedberg

Department of Linguistics
Simon Fraser University
hedberg@sfu.ca

Abstract

In this paper, we argue that in *it*-clefts as in *It was Ohno who won*, the cleft pronoun (*it*) and the cleft clause (*who won*) form a discontinuous syntactic constituent, and a semantic unit as a definite description, presenting arguments from Percus (1997) and Hedberg (2000). We propose a syntax of *it*-clefts using Tree-Local Multi-Component Tree Adjoining Grammar and a compositional semantics on the proposed syntax using Synchronous Tree Adjoining Grammar.

1 Introduction

The extant literature on the syntax of *it*-clefts, as in (1), can be classified into two main approaches. First, the cleft pronoun *it* is an expletive, and the cleft clause bears a direct syntactic or semantic relation to the clefted constituent, such as one of predication (Jespersen, 1937; Chomsky, 1977; Williams, 1980; Delin, 1989; Delahunty, 1982; Rochemont, 1986; Heggie, 1988; É. Kiss, 1998). Second, the cleft clause bears a direct syntactic or semantic relation to the cleft pronoun and is spelled-out after the clefted constituent through extraposition or by forming a discontinuous constituent with the cleft pronoun from the base-generated position at the end of the sentence (Jespersen, 1927; Akmajian, 1970; Emonds, 1976; Gundel, 1977; Wirth, 1978; Percus, 1997; Hedberg, 2000). Under this second approach, the cleft pronoun is not necessarily expletive but rather has a semantic function such as that of a definite article.

- (1) It was
 cleft pronoun + copula +
 OHNO [who won].
 clefted constituent + cleft clause

In this paper, we argue for a particular version of the second approach, in which the cleft pronoun and the cleft clause form a discontinuous syntactic constituent, and a semantic unit as a definite description. We propose a syntax of *it*-clefts using Tree-Local Multi-Component Tree Adjoining Grammar (MCTAG), and a compositional semantics on the proposed syntax using Synchronous Tree Adjoining Grammar (STAG). In section 2, we present arguments against the expletive approach, and in section 3, we provide arguments supporting the discontinuous constituent analysis. We present our TAG analysis in section 4 and extend our proposal to grammatical variations on *it*-clefts in section 5.

2 Arguments against the expletive approach

It has been shown in Hedberg (2000) that the cleft pronoun can be replaced with *this* or *that*, as in (2), depending on the discourse contextual interpretation of the cleft clause. The fact that the choice of the cleft pronoun is subject to pragmatic constraints indicates that the cleft pronoun cannot simply be an expletive element devoid of any semantic content.

- (2) a. This is not Iowa we're talking about.
 (Hedberg 2000, ex. 17)
 b. That's the French flag you see flying
 over there. (Hedberg 2000, ex. 20)

Although the details are different, many expletive analyses advocate for the position that the clefted constituent is syntactically associated with the gap in the cleft clause either directly through movement, or indirectly through co-indexation with an operator in the cleft clause. One thing that is common in all these analyses is that the cleft clause is not considered to have the internal structure of a restrictive relative clause. We point out

that the initial element in the cleft clause may be realized either as a *wh*-word (1) or as *that* (3a), or it may be absent altogether when the gap is not in the subject position (2, 3b). It may even be in the form of a genitive *wh*-word as in (3c). The cleft clause is thus a restrictive relative clause.

- (3) a. It was Ohno that won.
 b. It was Ohno Ahn beat.
 c. It was Ohno whose Dad cheered.

The cleft clause, however, does not relate to the clefted constituent in the way that a restrictive relative clause relates to its head noun, as first noted in Jespersen (1927). This is because the clefted constituent can be a proper noun, unlike a head noun modified by a restrictive relative clause, as illustrated in (4). This suggests that there is no syntactic link between the clefted constituent and the gap in the cleft clause.

- (4) * Ohno that won is an American.

3 A discontinuous constituent analysis

As pointed out in Percus (1997) and Hedberg (2000), *it*-clefts have existential and exhaustive presuppositions, just as definite descriptions do. The inference in (5c) associated with (5a) survives in the negative counterpart in (5b). This is exactly the way the presupposition associated with the definite description *the king of France* behaves: the presupposition spelled-out in (6c) survives in both the affirmative (6a) and the negative counterpart in (6b). Both authors argue that this parallelism between definite descriptions and *it*-clefts can be accounted for if the cleft pronoun and the cleft clause form a semantic unit, with *it* playing the role of the definite article and the cleft clause the descriptive component. What this translates to syntactically is that the cleft clause is a restrictive relative clause which is situated at the end of the sentence, forming a discontinuous constituent with the cleft pronoun.

- (5) a. It was Ohno who won.
 b. It was not Ohno who won.
 c. Someone won, and only one person won.
 (6) a. The king of France is bald.
 b. The king of France is not bald.

- c. There is one and only one king of France.

Percus (1997) further points out that *it*-clefts pattern with copular sentences containing definite description subjects with regard to anaphor binding. In the absence of c-command, an anaphor in the clefted constituent position can be bound by an antecedent inside the cleft clause, as shown in (7a). While we don't yet have an explanation for how this type of binding takes place, we follow Percus in noting that since copular sentences with definite description subjects also exhibit this pattern of binding, as shown in (7b), a uniform explanation for the two cases can be sought if the cleft pronoun and the cleft clause together form a definite description.

- (7) a. It was herself that Mary saw first.
 b. The one that Mary saw first was herself.

Under the discontinuous constituent analysis, *it*-clefts reduce to copular sentences, and therefore the observation that they can have equative and predicational interpretations (Ball 1978, DeClerck 1988, Hedberg 2000), the readings attested in copular sentences, follows. For instance, (5a) (repeated as (8a)) can be paraphrased as (8b), and corresponds to a typical equative sentence. And (9a) can be paraphrased as (9b), and corresponds to a typical predicational sentence. According to our analysis, (8a) will be assigned the semantic representation in (8c), and (9a) will be assigned the semantic representation in (9c).

- (8) a. It was Ohno who won.
 b. The one who won was Ohno.
 c. $THE_z [won(z)] [z = Ohno']$
 (9) a. It was a kid who beat John.
 b. The one who beat John was a kid.
 c. $THE_z [beat(z, John')] [kid(z)]$

4 Our TAG analysis

Inspired by work of Kroch and Joshi (1987) and Abeillé (1994) on discontinuous constituents resulting from extraposition, we propose a tree-local MCTAG analysis for the syntax of *it*-clefts. Crucially, we propose that the elementary trees for cleft pronoun and the cleft clause form a multi-component set, as in $\{(\alpha it), (\beta who_won)\}$ in Figure 1 and $\{(\alpha it), (\beta who_beat)\}$ in Figure 4.

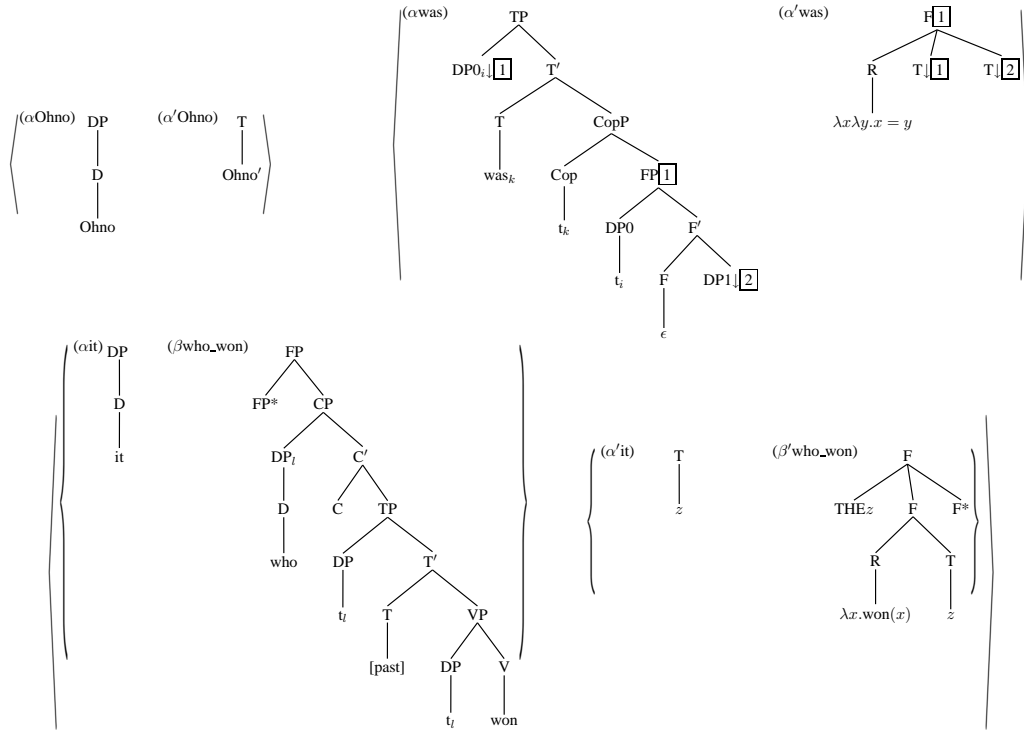


Figure 1: Syntactic and semantic elementary trees for *It was Ohno who won*

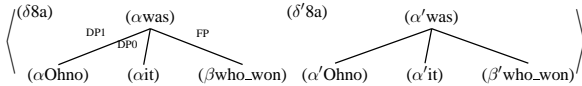


Figure 2: Syntactic and semantic derivation trees for *It was Ohno who won*

For the derivation of equative *it*-clefts as in (8a), we adopt the copular tree in (αwas) , a tree similar to the one proposed in Frank (2002) for copular sentences. In this tree, FP is a small clause of the copula from which the two DPs being equated originate. (8a) is derived by substituting (αit) into DP0 in (αwas) , adjoining $(\beta\text{who_won})$ into FP in (αwas) , and substituting (αOhno) into DP1 in (αwas) . The syntactic derivation tree and the derived tree for (8a) are given in $(\delta 8a)$ in Figure 2 and $(\gamma 8a)$ in Figure 3 respectively.

Postulating separate projections for the copula and the small clause can account for the fact that the clefted constituent and the cleft clause seem to form a constituent, as in (10ab) (from Hedberg 2000), and yet they can be separated by an adverbial phrase, as in (10c). In our analysis, (10ab) are possible because the bracketed parts are FPs. (10c) is possible because an adverbial phrase can adjoin onto FP or F', separating the clefted constituent and the cleft clause.

- (10) a. I said it should have been [Bill who negotiated the new contract], and it should have been.
 b. It must have been [Fred that kissed Mary] but [Bill that left with her].
 c. It was Kim, in my opinion, who won the race.

We propose to do compositional semantics using STAG as defined in Shieber (1994). In STAG, each syntactic elementary tree is paired with one or more semantic tree with links between matching nodes. A synchronous derivation proceeds by mapping a derivation tree from the syntax side to an isomorphic derivation tree in the semantics side, and is synchronized by the links specified in the elementary tree pairs. In the tree pairs given in Figure 1, the trees on the left side are syntactic elementary trees and the ones on the right side are semantic trees. In the semantic trees, F stands for formulas, R for predicates and T for terms. $(\alpha'\text{it})$ and $(\beta'\text{who_won})$ in the multi-component set in Figure 1 together define semantics of quantification, where the former contributes the argument variable and the latter the restriction and scope, and $(\alpha'\text{was})$ represents the semantics of equative sentences. The derivation tree for the semantics of (8a) is given in $(\delta'8a)$ in Figure 2, and the seman-

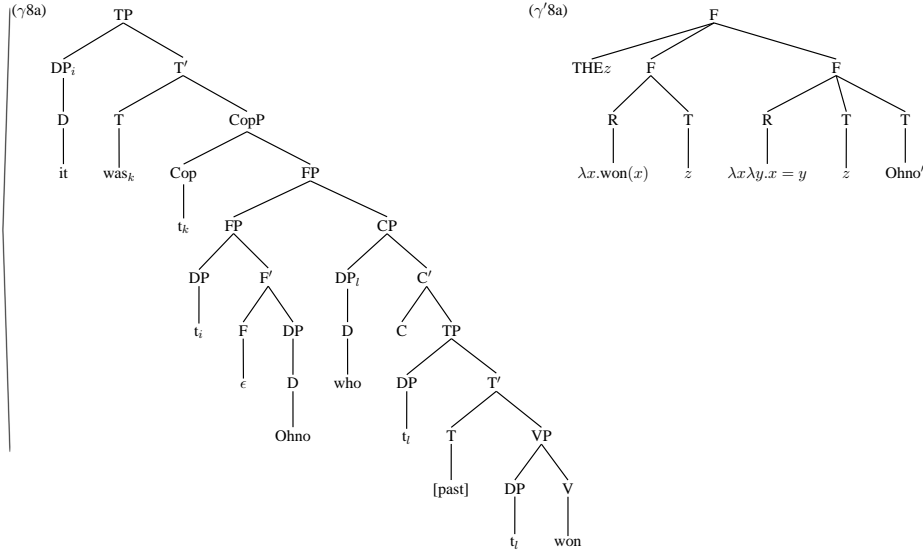


Figure 3: Syntactic and semantic derived trees for *It was Ohno who won*

tic derived tree is given in $(\gamma'8a)$ in Figure 3. Note that the semantic derivation tree in $(\delta'8a)$ is isomorphic to the syntactic one in $(\delta8a)$. The semantic derived tree in $(\gamma'8a)$ can be reduced to the formula in (11) after the application of λ -conversion.

$$(11) \quad \text{THE}z [\text{won}(z)] [z = \text{Ohno}']$$

For the derivation of predicational *it*-clefts as in (9a), we use the tree pairs in $\langle(\alpha\text{was_kid}), (\alpha'\text{was_kid})\rangle$, $\langle(\alpha\text{John}), (\alpha'\text{John})\rangle$, and $\langle\{(\alpha\text{it}), (\beta\text{who_beat})\}, \{(\alpha'\text{it}), (\beta'\text{who_beat})\}\rangle$ in Figure 4. The elementary tree in $(\alpha\text{was_kid})$ which represents a predicational copular sentence is similar to the one in (αwas) in that in both trees, the copula combines with a small clause FP. The important difference is that in $(\alpha\text{was_kid})$ the subject DP is an argument substitution site and the predicative DP (*a kid*) is lexicalized, whereas in (αwas) both the subject and the non-subject DPs are argument substitution sites. This difference is reflected in the semantic trees, as seen in $(\alpha'\text{was})$ in Figure 1 with two term nodes and $(\alpha'\text{was_kid})$ in Figure 4 with one term node. The syntactic and semantic derivation trees, which are isomorphic, are given in $\langle(\delta9a), (\delta'9a)\rangle$ in Figure 5, and the corresponding derived trees are given in $\langle(\gamma9a), (\gamma'9a)\rangle$ in Figure 6. The semantic derived tree in $(\gamma'9a)$ can be reduced to the formula in (12) after the application of λ -conversion.

$$(12) \quad \text{THE}z [\text{beat}(z, \text{John}')] [\text{kid}(z)]$$

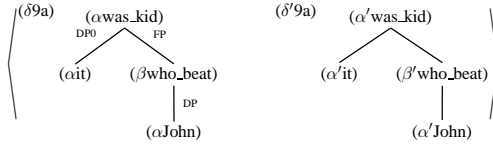


Figure 5: Syntactic and semantic derivation trees for *It was a kid who beat John*

5 Extensions

In this section, we extend the proposed syntactic analysis to grammatical variations on *it*-clefts: *wh*-extraction of the clefted constituent as in (13), unbounded dependency between the relative pronoun and its gap in the cleft clause as in (14), and coordination of the constituent containing the clefted constituent and the cleft clause as in (15).

(13) Who_j was it t_j who won?

(14) It was Ohno who_l the judges said t_l won.

(15) It was [Ohno who won] and [Kim who lost].

For the derivation of (13), the elementary trees in Figure 7 are required in addition to $\{(\alpha\text{it}), (\beta\text{who_won})\}$ in Figure 1. $(\alpha\text{who_was})$ represents the structure with the *wh*-extraction of the clefted constituent. Substituting (αwho) into DP1 and (αit) into DP0, and adjoining $(\beta\text{who_won})$ onto FP in $(\alpha\text{who_was})$, as in the derivation tree in $(\delta13)$, produces the derived tree in $(\gamma13)$ in Figure 8.

For the derivation of (14), the elementary trees in Figure 9 are required in addition to $\{(\alpha\text{it}),$

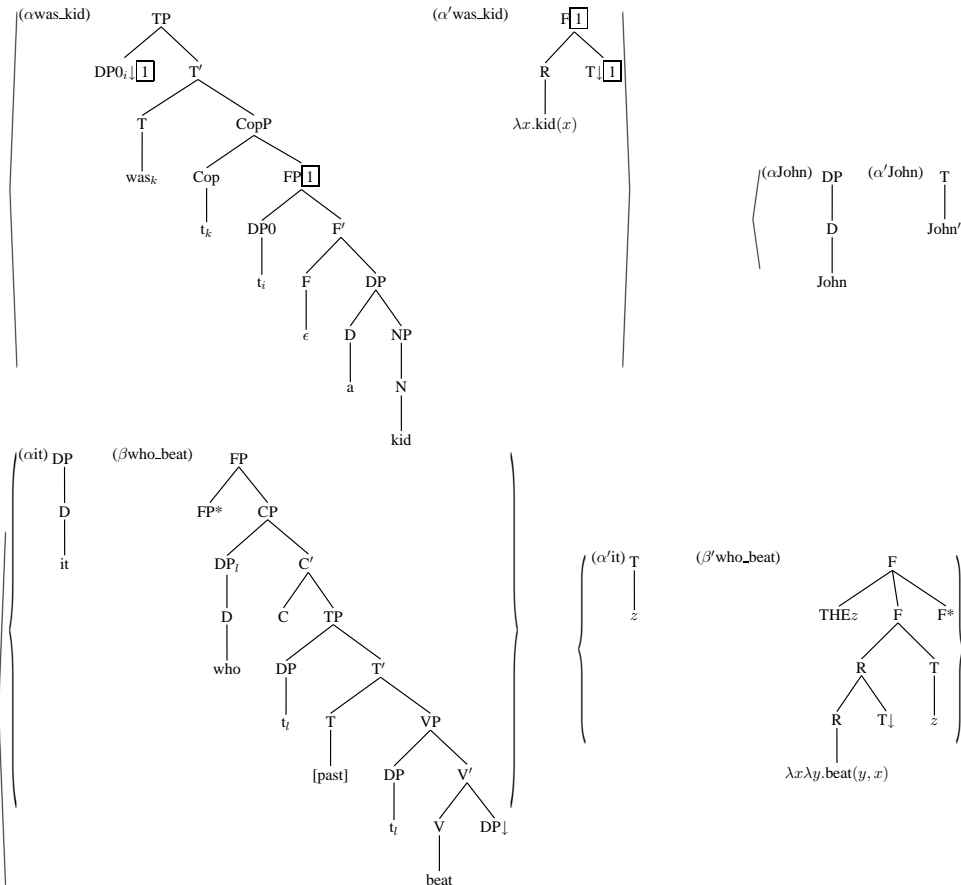


Figure 4: Syntactic and semantic elementary trees for *It was a kid who beat John*

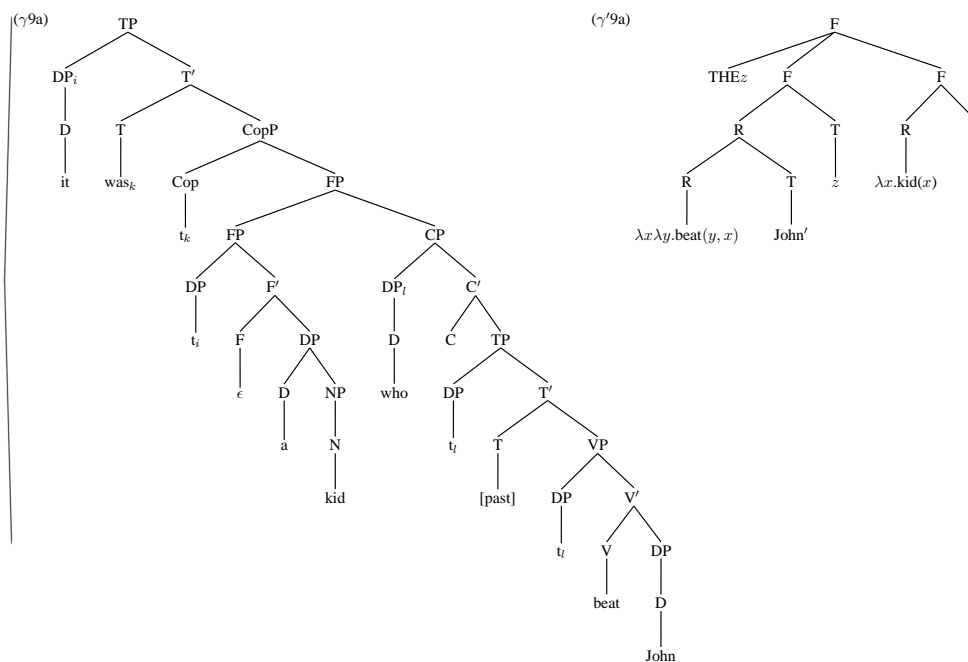


Figure 6: Syntactic and semantic derived trees for *It was a kid who beat John*

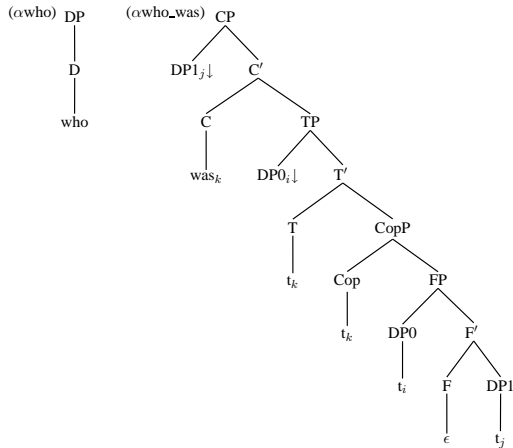


Figure 7: Syntactic elementary trees for *Who was it who won?*

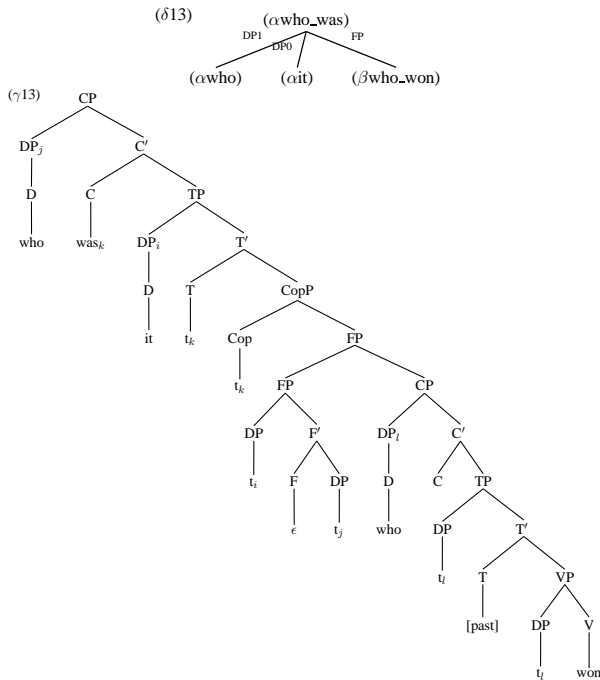


Figure 8: Derivation and derived trees for *Who was it who won?*

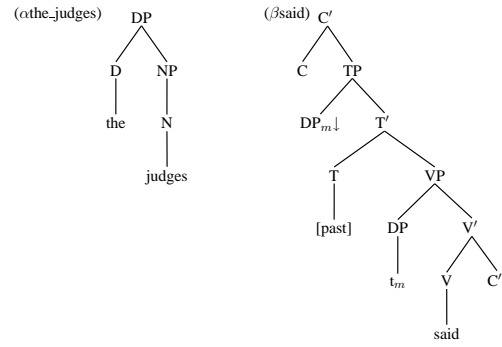


Figure 9: Syntactic elementary trees for *It was Ohno who the judges said won*

$(\beta\text{who_won})$ in Figure 1. Adjoining (βsaid) onto the C' node in $(\beta\text{who_won})$ has the effect of stretching the dependency between the relative pronoun *who* and its gap in the cleft clause. The derivation and the derived trees for (14) are given in Figure 10.

To handle the coordination of the constituent containing the clefted constituent and the cleft clause, as illustrated in (15), we propose to use Node Contraction and Conjoin proposed in Sarkar and Joshi (1996). Informally, Node Contraction takes two nodes of like categories and collapses them into a single node, and Conjoin coordinates the least nodes dominating the two contiguous strings. We use the conjunction tree in Figure 11 to apply Conjoin at FP.

Figure 12 contains the elementary tree anchoring equative *was*. We mark the nodes to be contracted with a box, and augment the name of the elementary tree with a set listing these contraction nodes. Thus, $(\alpha\text{was})\{DP_i, T, Cop\}$ means that DP_i , T and Cop nodes are marked for contraction in (αwas) elementary tree.

Composition of $(\alpha\text{was})\{DP_i, T, Cop\}$ tree in Figure 12 and another $(\alpha\text{was})\{DP_i, T, Cop\}$ tree with the conjunction tree in Figure 11, along with the substitution and adjoining of (αOhno) and an equivalent tree (αKim) anchoring *Kim*, $(\beta\text{who_won})$ and an equivalent tree $(\beta\text{who_lost})$ anchoring *lost*, and (αit) in appropriate places, yields the derived structure in Figure 13, where the contracted nodes get identified. In this structure, the DP hosting *it* is dominated by two TP nodes, T is dominated by two T' nodes and Cop is dominated by two CopP nodes. Thus, the derived structure produced by Conjoin and Node Contraction is a directed graph, not a tree.

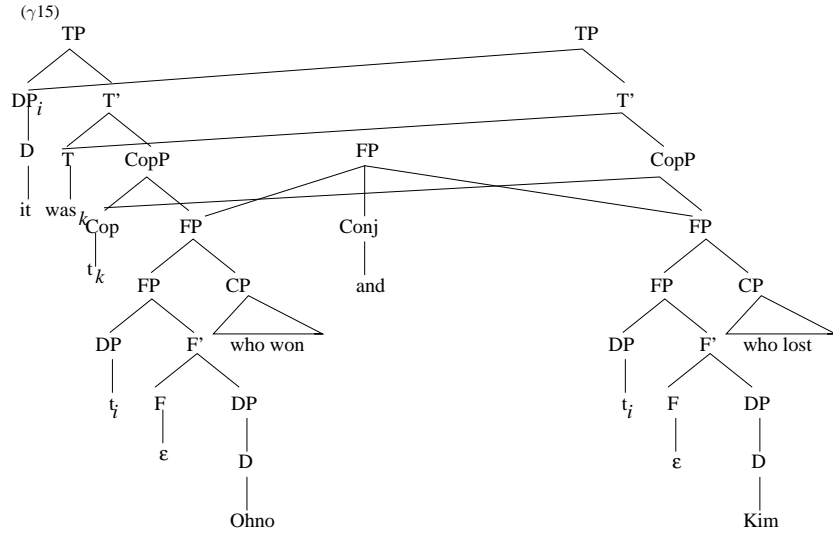


Figure 13: Derived structure for *It was Ohno who won and Kim who lost*

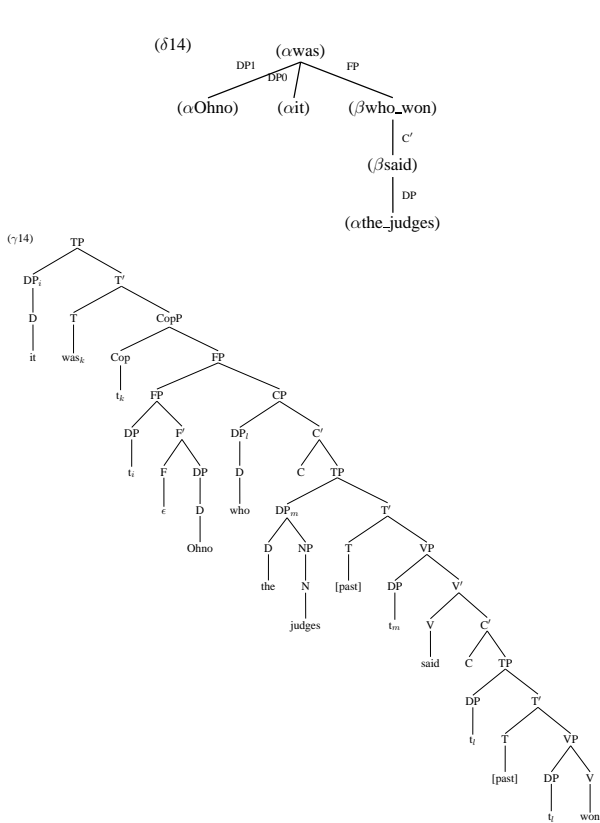


Figure 10: Derivation and derived trees for *It was Ohno who the judges said won*

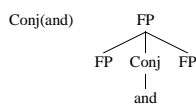


Figure 11: Elementary tree for conjunction

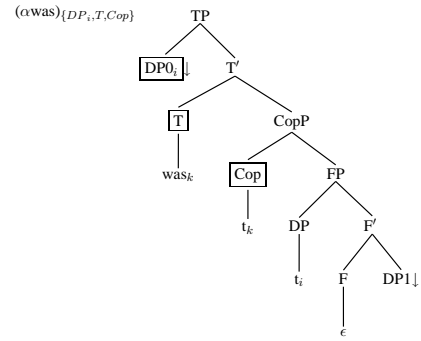


Figure 12: Elementary tree anchoring equative *was* with contraction nodes

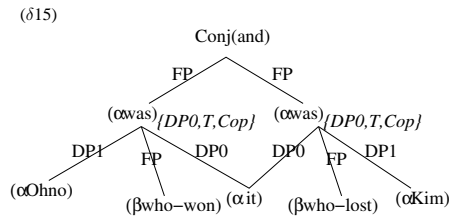


Figure 14: Derivation structure for *It was Ohno who won and Kim who lost*

The derivation structure for (15) is also a directed graph, as shown in Figure 14. (αit) is dominated by two $(\alpha was)_{\{DP_i, T, Cop\}}$ trees, indicating that it is being shared by the two $(\alpha was)_{\{DP_i, T, Cop\}}$ trees.

6 Conclusion

We have proposed a syntax and semantics of *it*-clefts, using tree-local MCTAG and STAG, and shown that the proposed syntactic analysis is ex-

tendable to handle various grammatical variations on *it*-clefts such as *wh*-extraction of the clefted constituent, unbounded dependency between the relative pronoun and its gap in the cleft clause and coordination of the constituent containing the clefted constituent and the cleft clause. In our TAG analysis of *it*-clefts, the cleft pronoun and the cleft clause bear a direct syntactic relation because the elementary trees for the two parts belong to a single multi-component set. They do not actually form a syntactic constituent in the derived tree, but as the elementary trees for the two belong to the same multi-component set, the intuition that they form a discontinuous constituent is captured. Further, the semantics of the two trees is defined as a definite quantified phrase, capturing the intuition that they form a semantic unit as a definite description.

Acknowledgment

We thank Anoop Sarkar and the three anonymous reviewers for their insightful comments.

References

- Ann Abeillé. 1994. Syntax or semantics? handling nonlocal dependencies with MCTAGs or Synchronous tags. *Computational Intelligence*, 10:471–485.
- Adrian Akmajian. 1970. On deriving cleft sentences from pseudo-cleft sentences. *Linguistic Inquiry*, 1:149–168.
- Noam Chomsky. 1977. On *wh*-movement. In P. W. Culicover, T. Wasow, and A. Akmajian, editors, *Formal Syntax*, pages 71–132. Academic Press, New York.
- Gerald P. Delahunty. 1982. *Topics in the syntax and semantics of English cleft sentences*. Indiana University Linguistics Club, Bloomington.
- Judy L. Delin. 1989. *Cleft constructions in discourse*. Ph.D. thesis, University of Edinburgh.
- Katalin É. Kiss. 1998. Identificatinoal focus versus information focus. *Language*, 74(245-273).
- Joseph E. Emonds. 1976. *A Transformational Approach to English Syntax*. Academic Press, New York.
- Robert Frank. 2002. *Phrase Structure Composition and Syntactic Dependencies*. MIT Press, Cambridge, MA.
- Jeanette K. Gundel. 1977. Where do cleft sentences come from? *Language*, 53:53–59.
- Nancy Hedberg. 2000. The referential status of clefts. *Language*, 76(4):891–920.
- Lorie A. Heggie. 1988. *The syntax of copular structures*. Ph.D. thesis, University of Southern California, Los Angeles.
- Otto Jespersen. 1927. *A Modern English Grammar*, volume 3. Allen and Unwin, London.
- Otto Jespersen. 1937. *Analytic Syntax*. Allen and Unwin, London.
- Anthony S. Kroch and Aravind K. Joshi. 1987. Analyzing extraposition in a Tree Adjoining Grammar. In G. Huck and A. Ojeda, editors, *Discontinuous Constituents*, volume 20 of *Syntax and Semantics*. Academic Press.
- Orin Percus. 1997. Prying open the cleft. In K. Kusumoto, editor, *Proceedings of the 27th Annual Meeting of the North East Linguistics Society*, pages 337–351. GLSA.
- Michael Rochemont. 1986. *Focus in Generative Grammar*. John Benjamins, Amsterdam.
- Anoop Sarkar and Aravind Joshi. 1996. Coordination in tree adjoining grammars: formalization and implementation. In *Proceedings of COLING'96*, pages 610–615, Copenhagen.
- Stuart Shieber. 1994. Restricting the weak-generative capacity of synchronous tree-adjoining grammars. *Computational Intelligence*, 10(4).
- Edwin Williams. 1980. Predication. *Linguistic Inquiry*, 11:203–238.
- Jessica R. Wirth. 1978. The derivation of cleft sentences in English. *Glossa*, 12(58-81).