Deriving Information Structure from Prosodically Marked Text with Lexicalized Tree Adjoining Grammars

Gann Bierner and Anoop Sarkar and Aravind Joshi
Dept. of Computer and Information Science
University of Pennsylvania
Philadelphia, PA 19104-6389
{gbierner, anoop, joshi}@linc.cis.upenn.edu

August 24, 1998

Abstract

This paper proposes a method for integrating intonation and information structure into the Lexicalized Tree Adjoining Grammar (LTAG) formalism. The method works fully within LTAG and requires no changes or additions to the basic formalism. From the existing CCG analysis, we denote boundary tones as lexical items and pitch accents as features of lexical items. We then show how prosodically marked text can be parsed to produce a derivation with the correct semantics and the appropriate information structure for the sentence. Although this paper is concerned with the recognition of prosodically marked text, the method described is also applicable to generation. This system has been implemented and tested using a wide-coverage LTAG grammar. The results in this paper also show how an account of intonational structure can be given in a lexicalized grammar with built-in constituencies in LTAG in contrast to lexical systems with flexible constituencies as in Combinatory Categorial Grammar (CCG).

Submission Type: Regular Paper

Topic Areas: L2. Syntax & parsing
L3. Semantics, pragmatics, cognition

Author of Record: Gann Bierner

Under consideration for other conferences (specify)? none

Deriving Information Structure from Prosodically Marked Text with Lexicalized Tree Adjoining Grammars

Abstract

This paper proposes a method for integrating intonation and information structure into the Lexicalized Tree Adjoining Grammar (LTAG) formalism. The method works fully within LTAG and requires no changes or additions to the basic formalism. From the existing CCG analysis, we denote boundary tones as lexical items and pitch accents as features of lexical items. We then show how prosodically marked text can be parsed to produce a derivation with the correct semantics and the appropriate information structure for the sentence. Although this paper is concerned with the recognition of prosodically marked text, the method described is also applicable to generation. This system has been implemented and tested using a wide-coverage LTAG grammar. The results in this paper also show how an account of intonational structure can be given in a lexicalized grammar with built-in constituencies in LTAG in contrast to lexical systems with flexible constituencies as in Combinatory Categorial Grammar (CCG).

1 Introduction

The correspondence between intonation and information structure has been explored in (Selkirk, 1984; Pierrehumbert and Hirschberg, 1990; Steedman, 1991). Such a correspondence can be exploited in several situations; for example, given some discourse knowledge, it can enhance text-to-speech systems by allowing them to directly produce reasonable intonation without applying heuristics, e.g. (Prevost and Steedman, 1994). Conversely, systems that attempt to infer the discourse function of utterances from speech are greatly improved by considering prosody.

Consequently, representing prosody within a grammar formalism offers an advantage in any work related to speech understanding or content-to-speech planning. With this representation the interpretation and generation of prosodically marked text can be handled in the derivation process of the parser/generator. This paper focuses on the derivation of information structure from prosodically marked text, but our work also applies to generation of prosodically marked utterances given the information structure.

(Steedman, 1991) has pointed out the correspondence between intonational phrases and the constituent structure of coordinate sentences. He demonstrated in CCG, a formalism with flexible constituencies, that each prosodic consituent corresponds to a syntactic constituent. We present a similar account in LTAG (Joshi, 1987; Joshi and Schabes, 1992), a formalism without flexible constituency. We show how prosodic markers can be easily integrated into LTAG (with no changes to the basic formalism) much in the same way as the (Steedman, 1997) account in CCG so that the appropriate information structure is constructed in the derivation of an utterance. When we consider the treatment of coordination in LTAG, we show that our analysis captures the relation between intonational phrases and constituency in coordinate sentences.

2 Intonation and Information Structure

Intonation:
Intonation can be represented by assigning particular melodies, consisting of boundary tones and pitch accents, to an utterance. We will use Pierrehumbert's (Pierrehumbert and Hirschberg, 1990) notation for prosody throughout the paper. The prosodic building blocks are

H and **L** which respectively conform to the intuitive notion of *high* and *low* pitch.

Pitch accents consist of **L**s and **H**s connected with '+' signs, with a '*' somewhere in the string. The '*' indicates that the pitch will culminate at the stressed syllable of the marked word. For instance, the word *computer* marked by L+H* would begin at a low pitch and would rise until culminating within the middle (stressed) syllable, before falling off again.

Boundary tones such as LL% and LH§ indicate the rising or falling at the end of intonational phrases. Boundary tones ending with % are utterance medial while boundary tones ending in § indicate the end of an utterance.

Information Structure:
Prosodic melodies divide an utterance into prosodic constituents that correspond to the utterance's *theme* and *rheme.* The theme of an utterance is the open proposition that is currently under scrutiny. This proposition may be satisfied by any number of alternatives. The rheme is the portion of the utterance that provides a specific entity that satisfies the proposition. More formally, (Steedman, 1997) defines theme and rheme as in (1).

(1) Theme: The alternative set associated with an utterance.
Rheme: Restricts the alternative set.

(2) Who likes spam?
(Mary LL%) (likes spam LH§)
H* L+H*

Theme: $\lambda x.likes(x, spam')$
Rheme: $mary'$

As we can see in (2), the theme represents all the people who *like spam* and the rheme restricts this set of people to *Mary.* The theme and rheme taken together form the utterance's *information structure.* Our goal is to correctly derive the information structure of an utterance to update a discourse model. For example, the discourse model could be a continuously updated database of some interactive application such as tutorial program.

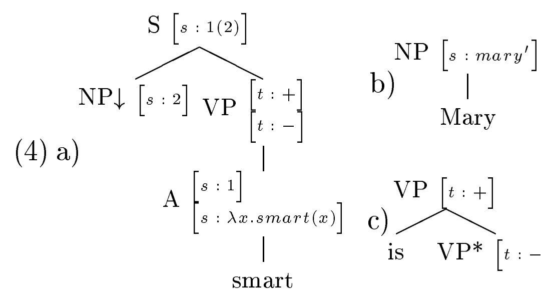
Acording to (Steedman, 1997), particular pitch accents are associated with theme (θ) or rheme (ρ), (3).

(3)
L+H*, L*+H	:=	θ
H*, L*	:=	ρ

Boundary tones split an utterance into prosodic constituents which may either be themes or rhemes as indicated by the pitch accents. Consequently, lexical items within a constituent can only be marked with pitch accents associated with theme or rheme, but not both.

3 Information Structure and LTAG

In this section we describe our approach in detail. We start with a brief introduction to the LTAG formalism[1]. We use a unification based LTAG where each node in a tree is embellished with feature structures. LTAG is a tree-rewriting system where the grammar consists of a set of trees called elementary trees, e.g. those shown in (4).

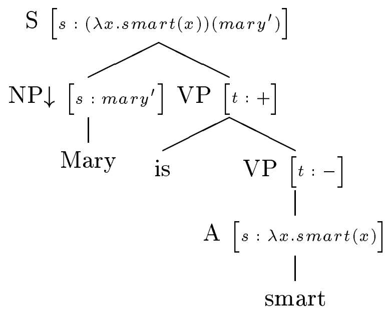


Trees can be rewritten using two operations: substitution and adjunction. For example, the tree (4b) substitutes into the NP node in tree (4a), while the tree (4c) adjoins into the VP node of (4a). The VP node marked with a '*' is referred to as the *foot node.* The feature structure on the NP node of (4a) restricts the kind of trees that can substitute there. An internal node such as the VP in (4a) has two feature structures associated with it, called the *top* and *bottom* features as the internal node is replaced by two nodes after adjunction[2]. At the end of a derivation, the top and bottom features in each node are unified. In (4) this forces the adjunc-

[1]For a more detailed introduction to LTAG and the notion of derivation in an LTAG, see (Joshi, 1987; Joshi and Schabes, 1992)

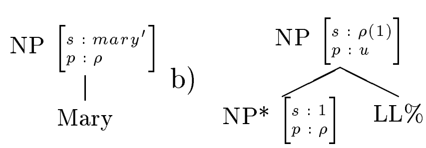
[2]To simplify the exposition, we will not show the top feature structure if it contains no relevant features.

tion of the tree anchored by *is* into the one anchored by *smart* at the VP node to give us the valid derivation in (5).

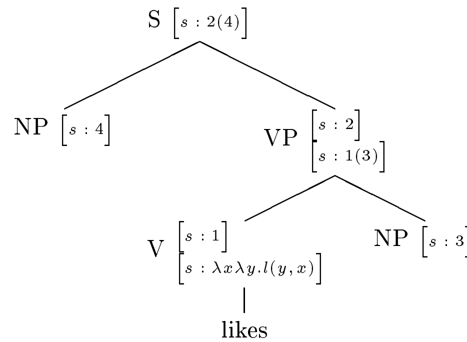
(5) 

We use features in an elegant way to assemble prosodic constituents. Lexical items marked by pitch accents are represented with prosodic features corresponding to (3) and may be instatiated to θ, ρ, or u. θ and ρ correspond to theme and rheme and u denotes a phrase that has already been marked by a boundary tone. So, the lexical item *Mary* marked by an H* pitch accent would produce a tree with a ρ prosodic feature as in (6a). More specifically, as in (Steedman, 1997), ρ and θ are identity functions that when applied, update the database associated with the discourse model with a rheme and theme, respectively.

As noted in Sec. 2, boundary tones separate utterances into prosodic constituents. We follow (Steedman, 1997) in modeling boundary tones as lexical items which select their own trees in the grammar. These lexicalized trees combine with the other items in the lexicon by adjoining to the appropriate node and by updating their semantics to theme or rheme depending on their pitch accent. The subtree of the adjoined node becomes a prosodic constituent, and u is used to denote that a constituent cannot accept further prosodic marks. For example, an LL% boundary tone appearing after the rhematically marked *Mary* would be the tree in (6b). Unlike CCG, we require that the prosodic feature of the foot node be either ρ or θ so that boundary tones do not adjoin twice to the same position.

(6) a) b) 

The key to integrating intonation and information structure with LTAG is the incremental construction of the semantics in an LTAG verb tree[3]. The verb node of the tree contains the proposition associated with the verb. For the verb *likes*, the proposition would be $\lambda x \lambda y.l(y,x)$. The semantics for each verb argument in the tree is applied to the open proposition at the verb argument's parent node. Thus, the *trunk* of the tree (the path from the root to the anchor) contains an incremental derivation of the semantics, (7).

(7) 

In (8), we show schematically how the prosodic features are distributed over the verb tree. In order to separate the tree into prosodic phrases, one boundary tone must adjoin to a node along the trunk of the tree, and the other boundary tones must adjoin to the verb arguments in the tree above this adjunction site (the super tree). This separates the tree into prosodic constituents, the subtree below the adjunction site on the trunk and the verb arguments in the supertree. The obligatory adjunction of the boundary tone is ensured by using the top and bottom prosodic features on each internal node. For a node along the trunk of the tree, the bottom prosodic feature is coindexed with the prosodic feature of its verb argument. Its top prosodic feature is coindexed with the verb argument of its parent node, as with VP in (8).

Since the top and bottom features of a particular node must unify at the end of the derivation, the features must both be ρ or θ (in the subtree), both be u (in the supertree), or be sep-

[3]The construction of the semantic features in the verb tree has interesting links with the treatment of coordination in LTAG. See the discussion in Sec. 6.

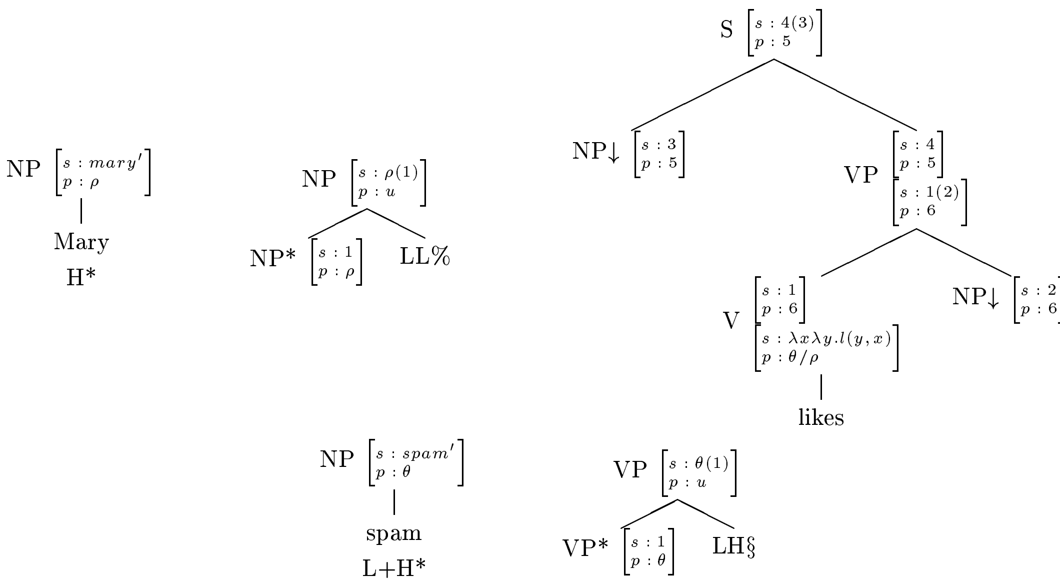
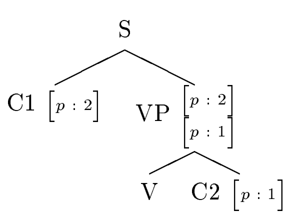


Figure 1: Elementary Trees for *(Mary) (likes spam).*

arated by the adjunction of a boundary tone. Since all verb arguments in the tree are connected through this chain of features, the features within a prosodic constituent can either be ρ or θ but not both. Also, it should be noted that these trees can still be used with text unannotated by prosody. The added features just remain uninstantiated and trivially unify. They do not interfere, and the derivation proceeds similarly to the derivation in (5).

(8) 

4 Examples

Consider a derivation of (2) using the trees in Fig. 1.

The LL% tree adjoins to the *Mary* tree. This forms a derived tree whose root prosodic feature is instantiated to u. When this derived tree substitutes into the *likes* tree, the top prosodic feature of the VP node is instantiated to u. This creates a potential feature clash. The final boundary tone adjoins to the VP node to split the conflicting features. If we had a boundary tone that could adjoin to the NP node, it would not be used because it instantiates the top prosody feature of V to u, causing another feature conflict. Since no boundary tones remain to adjoin to this node and separate the features, the conflict cannot be resolved.

Therefore, the correct tree is derived in Fig. 2. The root node contains the unevaluated form of the semantics in lambda notation. When evaluated, $\theta((\lambda x \lambda y.l(y,x))(spam'))(\rho(mary'))$ yields the appropriate semantics $l(mary', spam')$. It also updates the theme as $\lambda y.l(y, spam')$ and rheme as $mary'$, which gives us a correct derivation for (2).

Another example is shown in (9). Fig. 3 shows the elementary trees for this sentence.

(9) What does Mary like?
(Mary likes LH%) (spam LL§)
L+H* H*

Theme: $\lambda x.likes(mary', x)$

Rheme: $spam'$

Notice that the *likes* tree for this example contains an extraposed object. As stated in Sec. 3, the tree must contain an incremental derivation of the semantics along the trunk to provide prosodic constituents. For this example, we require $l(mary', x)$ to appear somewhere in the derivation, and this does not occur in the

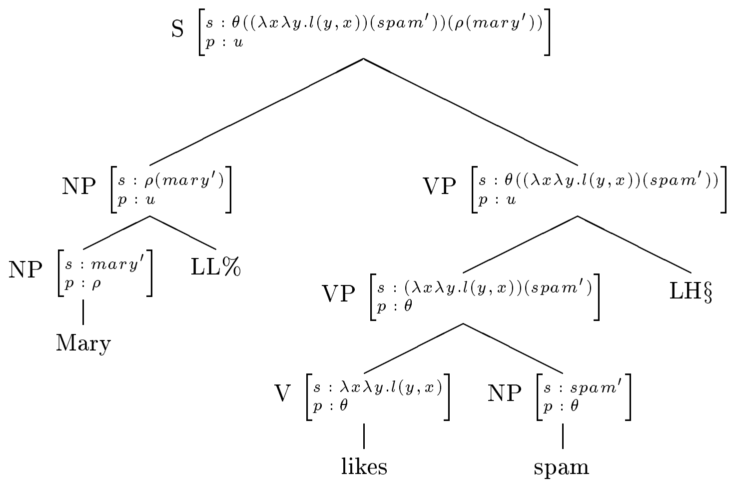


Figure 2: Derived Tree for *(Mary) (likes spam).*

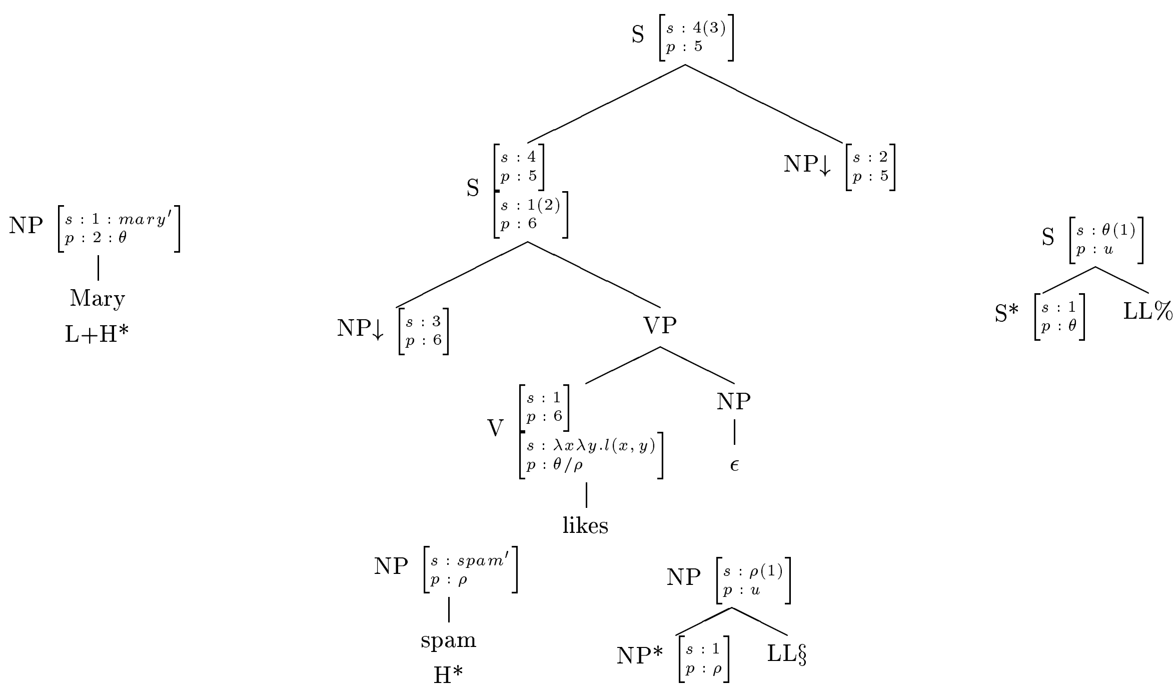


Figure 3: Elementary Trees for *(Mary likes) (spam).*

original *likes* tree in Fig. 1. We could modify the feature structures of the original tree to reflect the new derivation, but it would be odd not to have the semantics of the verb argument apply at its parent. Instead we chose to build a new tree that correctly reflects the new semantics. This tree is also motivated in other ways as it is used in the LTAG treatment of coordination. We point out in Sec. 6 that the prosodic constituents created in our approach match exactly the constituents created in cases of coordination. This holds even of examples of traditional non-constituents in both our analysis of prosody and the existing analysis of coordination in TAGs. With features, it is trivial to ensure that this tree will only be used in the case of prosody or coordination and not for normal transitive constructions.

The elementary trees then combine similarly to the first example to finally produce the tree in Fig. 4 with the semantics: $\theta((\lambda x \lambda y.l(x,y))(mary'))(\rho(spam'))$. When evaluated, this produces appropriate theme and rheme, as in (9).

5 Evaluation

The examples in Sec. 4 as well as other crucial constructions[4] have been implemented in the XTAG system (Group, 1995), a wide coverage grammar and parser for English. The implementation confirms important features of our formalism.

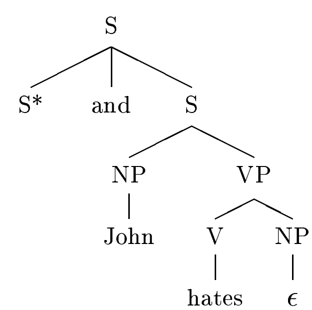
First, the formalism works as expected for our examples by producing the appropriate semantics and information structure. Second, no extraneous derivations occur. That is, because of the constraints imposed by the prosodic features in the verb tree, the elementary trees for the prosodically marked sentences do not combine in unexpected ways. Finally, also due to the prosodic features, sentences with both theme and rheme pitch accents in a single prosodic constituent are rejected, as we might intuitively expect.

[4]These include small clauses, sentential complements, and relative clauses.

6 Discussion

Much of this work has been inspired by the developments in categorial grammars, in particular in the combinatory categorial grammars (CCG) (Steedman, 1997). In a sense this is not surprising as LTAG and CCG are known to be (under certain conditions) weakly equivalent. However, it is not obvious either as they are not strongly equivalent (Joshi et al., 1991). LTAG is a tree-rewriting system and has built-in constituency. In CCG the notion of constituency is very flexible. In fact this very flexibility is what enables it to provide novel accounts of coordination and intonational structure. In some of our earlier work we have shown how the essential ideas about coordination in CCG can be articulated in the LTAG framework of predefined constituency (Joshi and Schabes, 1991; Anonymous, 1996). The results in this paper can be interpreted as carrying out a similar program in the domain of intonational structure.

(10)



As in CCG there is direct relation between the prosodic constituents imposed by our approach and constituents imposed by coordination. Take for example, "Mary likes and John hates spam." The LTAG analysis for this sentence uses the tree for *likes* in Fig. 3 with the structure for *hates* in (10)[5]. The tree in (10) adjoins into the tree in Fig. 3. The prosodic constituent in the tree is identical to the constituent in the coordinate sentence. The reason for using the extraposed tree for prosody is linked strongly to the fact that the right semantics has to be assigned. This is parallel to the case in coordination where the extraposition is

[5]Strictly speaking, the tree in (10) has an associated structure for the extraposed object of *hates*. This structure is shared when (10) adjoins into the *likes* tree. The details of this can be found in (Anonymous, 1996).

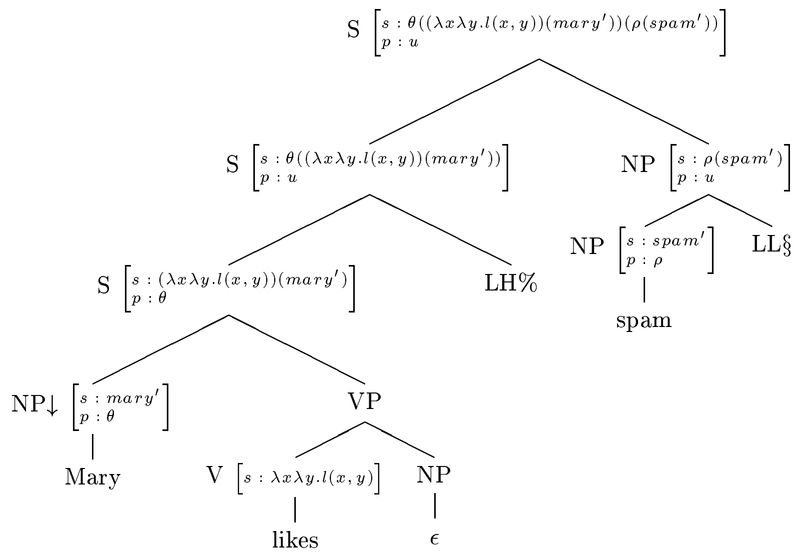


Figure 4: Derived Tree for *(Mary likes) (spam).*

crucial in the sharing of the object structure between the two verbs: also a case of the correct semantics being assigned.

7 Conclusion

In this paper we have shown a method of integrating intonation and information structure with the LTAG grammar formalism. We demonstrate how prosodically marked text can be parsed to produce appropriate semantics and information structure. We also show how the relationship between intonational phrases and constituents in coordinate sentences can be captured in LTAG.

References

Anonymous. 1996. Coordination in tree adjoining grammars: Formalization and implementation. In *Proceedings*.

XTAG Research Group. 1995. A lexicalized tree adjoining grammar for english. Technical report, University of Pennsylvania. IRCS Technical Report IRCS-95-03, ftp://ftp.cis.upenn.edu/pub/ircs/tr/95-03.ps.Z.

A. Joshi and Y. Schabes. 1991. Fixed and flexible phrase structure: Coordination in Tree Adjoining Grammar. In *Proc. 5th Int'l Workshop on Natural Language Generation*.

A. Joshi and Y. Schabes. 1992. Tree-adjoining grammar and lexicalized grammars. In M. Nivat and A. Podelski, editors, *Tree automata and languages*, pages 409–431. Elsevier Science.

A. Joshi, K. Vijay-Shanker, and D. Weir. 1991. The convergence of mildly context-sensitive grammatical formalisms. In P. Sells, S. Shieber, and T. Wasow, editors, *Foundational issues in natural language processing*, pages 31–81. MIT Press.

A. Joshi. 1987. An introduction to tree-adjoining grammars. In Alexis Manaster-Ramer, editor, *Mathematics of language*, pages 87–115. John Benjamins.

J. Pierrehumbert and J. Hirschberg. 1990. The meaning of intonational contours in the interpretation of discourse. In *Intentions in Communication*, pages 271–312.

S. Prevost and M. Steedman. 1994. Specifying intonation from context for speech synthesis. *Speech Communication*, pages 139–153.

E. Selkirk. 1984. *Phonology and Syntax*. MIT Press.

M. Steedman. 1991. Structure and intonation. *Language*, 67:260–296.

M. Steedman. 1997. Information structure and the syntax-phonology interface. Under review, http://www.cis.upenn.edu/~steedman/papers.html.