Using APML to Specify Intonation.*

Mark Steedman
University of Edinburgh
steedman@informatics.ed.ac.uk

Draft 1.1, February 13, 2004

Abstract

This is a short introductory guide to using APML markup to annotate spoken texts and to specify intonation for the FESTIVAL open-source speech synthesizer.

1 Introduction

APML is an XML-based markup language for specifying turn-taking, performative, affective, and intonational aspects of text for speech synthesisis and facial animation (de Carolis et al. 2004).

The present memo describes the intonational control of speech synthesis using APML for these purposes. The paper begins by sketching a compositional Information Structural Semantics for the intonational tones, as suggested by Liberman (1975), along somewhat similar lines to Pierrehumbert and Hirschberg (1990) and Steedman (2004). The tones are of two kinds: *pitch accents*, which are perceived as word-level stress, accent or emphasis, and realized as the APML elements labeled as `emphasis`; and *boundaries*, perceived as prosodic phrase-final rising pitch and/or lengthening, and realized as APML `boundary` elements. This section is included mainly to explain why intonation in English is assigned as it is by speakers of the language, and might be skimmed on a first reading.

The paper then introduces APML, and turns to the question of how the working Language Technician can apply intonational annotation correctly, either as an annotator of human speech, or as a programmer defining a generator or realizer for intonationally appropriate speech.

2 Some Theory: Tones and Information Structure

It is standard to assume, following Bolinger (1958, 1961) and Halliday (1963, 1967a,b), that pitch-accents, high or low, simple or compound, are in the first place properties of the words that they fall on, and that they mark the interpretations of those words as contributing to the distinction between the speaker's actual utterance and other things that they might be expected to have said in the context to hand, as in the "Alternative Semantics" of Kartunnen (1976),

*The work was supported in part by EU FET grant MAGICSTER and the Edinburgh-Stanford Link grant "Sounds of Discourse.".

Karttunen and Peters (1979), Rooth (1985, 1992), and Büring (1997a,b).[1] In this sense, all pitch accents are *contrastive*. For example, in response to the question "Which finger did he bite?", the word that contributes to distinguishing the following answer from other possible answers via reference is the deictic "this", so the following intonation is appropriate.[2]

(1) He bit THIS finger .
H* LL%

It is important to be clear from the start that the set of alternative utterances from which the actual utterance is distinguished by the tune is in no sense the set of all possible utterances appropriate to this context, a set which includes infinitely many things like "Mind your own business," "That was no finger," "What are you talking about?" and "Lovely weather we're having." Rather, the presupposed set of alternative utterances is *accommodated* by the hearer in the sense of Lewis (1979) and Thomason (1990), like any speaker presupposition that is not actually inconsistent with their beliefs. This does not imply that such alternative sets are confined to things that have been mentioned, or that they are mentally enumerated by the participants—or indeed that they are even finite.

In terms of Halliday's given/new distinction pitch-accents are markers of "new" information, although the words that receive pitch-accents may have been recently mentioned, and it might be better to call them markers of "not given" information. That seems a little cumbersome, so I will borrow the notion of "kontrast" from Vallduví and Vilkuna 1998 for this property of English words bearing pitch-accents.[3]

Steedman 2000 and Vallduví and Vilkuna 1998 argue that there are just two independent semantic binary-valued dimensions along which the literal meanings of the various pitch-accent types are further distinguished. The first of these dimensions has been identified in the literature under various names, and distinguishes between what I'll continue to call "theme" and "rheme" components of the utterance, using these terms in the sense of Bolinger (1958, 1961) rather than Halliday. Theme can be thought of informally as the part of the sentence corresponding to a question or topic that is presupposed by the speaker, and rheme is the part of the utterance that constitutes the speaker's novel contribution on that question or topic. For example, in response to the question "I know which finger the FROG bit. But which finger did FISH bite?", the word that contributes to distinguishing the theme int the following answer from other possible themes is the pronoun "He", so the following verson of (1), in which the tune L+H* LH% marks a theme, linking the answer to the question, is appropriate.

(2) HE bit THIS finger .
L+H* LH% H* LL%

[1] The term "pitch-accent" is here restricted to what Ladd (1996) calls "primary" pitch-accents, sometimes called "nuclear" pitch accents (although there may be more than one in a sentence). Ladd follows Halliday in distinguishing primary accents from certain other accents that arise from the interaction of lexical stress with metrical the metrical grid. While there is still no objective measure to distinguish the two varieties, it is the primary accents that are perceived as emphatic or contrastive.

[2] The notation for tunes is Pierrehumbert's, see Pierrehumbert and Hirschberg 1990 for details including characteristic pitch-contours.

[3] In Steedman 2000 and earlier work I called this property "focus", following the "narrow" sense of Selkirk (1984). However, this term invites confusion. Ladd (1980) defi ned another "broad" sense of the term, which is the one intended by Hajičová and Sgall (1988) and Vallduví (1990), which is closer to the term "rheme" as used in the present system, and in Steedman 2000 and Vallduví and Vilkuna 1998.

It is important to recognize that this utterance is also possible in the context that was used for example (1). In that case, the intonationally marked theme will give rise to a further *implication* about the context, namely that that it is consistent with another possible theme concerning someone *else* biting. This implicature will cause the hearer to update his or her model of the domain of discourse, to (in the sense defined earlier) *accomodate* another (underspecified) individual. This observation, discussed further below, should make it clear that the notion of theme differs from that of topic as defined by, for example, Gundel (1974); Gundel and Fretheim (2001) in being *speaker-defined* rather than text-based.

A great deal of the huge and ramifying literature on information structure can be summarized as distinguishing two dimensions corresponding to the background/kontrast and theme/rheme distinctions, although the consensus has tended to be obscured by the very different nomenclatures that have been applied. (See discussion by Kruijff-Korbayová and Steedman (2001), which summarizes the terminology and its lines of descent, along with some contiguous semantic influences.)

However, there is a further dimension of discourse meaning along which the pitch-accent types are distinguished which has not usually been identified in this literature. It concerns whether or not the particular theme or rheme to hand is mutually agreed–that is, uncontentious. This notion is related to various notions of Mutual Belief or Common Ground proposed by Lewis (1969), Chafe (1974), Cohen (1978), Clark and Marshall (1981) and Clark (1996).[4]

All three of these components of meaning are projected by the process of grammatical derivation from the words that carry the pitch-accent to the prosodic phrase corresponding to these information units, by a grammatical mechanism described in Steedman 2000.

I'll also try to argue that the intonational boundaries such as those sometimes referred to as "continuation rises," which delimit the prosodic phrase, fall into two classes respectively distinguishing the speaker or the hearer as responsible for, "owning", or (in terms of the related accounts of Gussenhoven (1983, p.201) and Gunlogson (2001, 2002)) **committed to**, the corresponding information unit.

I'll assume that the speaker's knowledge can be thought of as a database or set of propositions in a logic (second-order, since themes etc. may be functions), divided into two subdomains, namely: a set S of information units that the speaker claims to be committed to, and a set H of information units which the speaker claims the hearer to be committed to. Information units are further distinguished on a dimension $\pm$*AGREED* according to whether the speaker claims them to be uncontentious or contentious. The set of $+$*AGREED* information units is not merely the intersection of S and H: the speaker may *attribute* uncontentiousness to an information unit and responsibility for it to the hearer whilst knowing that in fact they do not regard themselves as so committed. In Steedman 2000, S and H are treated as modalities $[S]$ and $[H]$ of a modal logic, and Stone (1998) has proposed a similar modality for mutual belief. In the present paper we will combine the feature $\pm$*AGREED* with the speaker/hearer modalities, writing it as a superscript $\pm$, as in $[H^{+}]$.

These classifications can be set out diagrammatically as in the tables 1 and 2, in which θ signifies theme, ρ signifies rheme, $+$ indicates $+$*AGREED*, $-$ indicates $-$*AGREED*, and $[S]$ and $[H]$ respectively denote speaker and hearer commitment. If a theme or a rheme is marked as

[4]Hobbs (1990), who proposes a very different revision of Pierrehumbert and Hirschberg (1990) to the present one, also gives a central role to Mutual Belief.

	+	−
θ	L+H*	L*+H
ρ	H*, (H*+L)	L*, (H+L*)

Table 1: The Meanings of the Pitch-accents

$[S]$	L, LL%, HL%
$[H]$	H, HH%, LH%

Table 2: The Meanings of the Boundaries

agreed, then it's in *AGREED*, whoever is explicitly claimed to be committed to it. If it is not so marked, then it is not in *AGREED*, *even if speaker and hearer in fact both believe it*. This last possibility arises because *H* is only the speaker's attribution of commitment to the hearer, not the hearer's actual belief. It follows that a theme or rheme may be believed by the speaker, and asserted by the speaker to be something that the hearer is committed to, without the hearer's actually agreeing to it. We will come to a case of this later on.

These elements of meaning are very abstract, and their use by speakers to convey the relation of what they are saying to the surrounding discourse are often extremely indirect, the conveyed meaning depending on the process of accomodation by the hearer of the speakers claims about information units and the participants' commitment and agreement to them. This complexity, coupled with the lack of any fully satisfactory instrumental definition of the tones themselves, makes the task of the annotator in particular extremely difficult. However there are some rules of thumb which can help with this problem, and which if applied to the process of "content to speech" synthesis at least, seem to give good results (Prevost 1995; Prevost and Steedman 1994, Cassell et al. 1994). We will take the tasks of annotation and APML-based synthesis in turn.

3 Annotating Intonation

The first thing to know about English tones is that their category is not at all perceptually obvious to the untrained ear. Even the distinction between H* and L* is easy to mistake for the novice, and some training is usually needed.

The best presentation of the Pierrehumbert tone system is in Pierrehumbert and Hirschberg 1990, and in particular the reader is directed to figures 14.13 and 14.14, p.281, in which schematic f_0 contours are shown for all pitch accents in combination with all following boundaries. This system formed one part of the basis for the ToBI annotation scheme (Silverman et al. 1992), which is the format in which most machine-readable corpora have been produced.

3.1 The Nature of ToBI Annotation

Tobi annotated-corpora need to be treated with some caution. Because it was devised in the Bolingerian theoretical tradition of Pierrehumbert, Tobi annotation does not distinguish the prominent, "primary," kontrast-related pitch-accents from "lexical" pitch accents. While there is

no formal or instrumental definition of this prominence distinction, it is important, because only primary accents are relevant to information structure as defined here. Moreover, though ToBI is the most completely worked out and linguistically well-found annotation scheme we have, inter-annotator reliability is quite low on some of the distinctions that are most relevant to information structure—particularly the distinction between H*, L+H*, and L*+H accents (Pitrelli, Beckman and Hirschberg 1994, Taylor 2000). The latter paper shows the extent of the overlap in the space of TILT parameters via Figure 1, based on a ToBI annotated corpus.

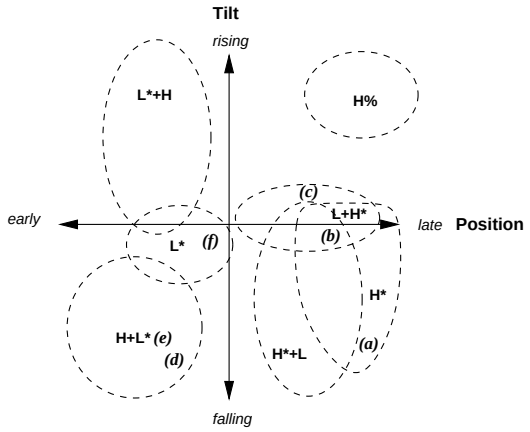

Figure 1: Examples of Tilt Categorization of results of TOBI hand-annotation(from Taylor 2000)

The main reason for this interannotator unreliability is probably that the ToBI instructions encourage the use of H* as a default, and define pitch accent contours solely in terms of shape. The overall shape of L+H* and some kinds of H* (those preceded by low pitch) can be very similar. While the parameters that distinguish them are a matter of current research, two very promising candidates are: alignment of the start of the rise with the syllable/vowel onset; and relative pitch heights. the rise of an L+H* accents is delayed until the vowel onset, while an H* may rise at any earlier point, or indeed be high from the start. L+H* accents are also typically lower than H* (Ladd and Schepman 2003; Calhoun 2004; cf. Pierrehumbert and Hirschberg 1990:281).

To successfully do intonational annotation for extraction or realization of information structure in the face of this uncertainty, it helps to keep the following principles in mind.

1. Emphases are properties of lexical items—usually, single words. Think of them as equivalent to underlining or italicising. Boundaries are string items in their own right. Think of them as punctuation.

2. It follows that decisions about tone-categories can only be taken in the knowledge of the segment- syllable- and word-level structure of the speech. Words tell you where lexical stresses are predicted, and primary pitch accents always align with lexical stress (except in rare cases of lexical contrast). Word boundaries are similarly important to defining possible positions for boundary tones. For example, it is in many cases only by access to

the transcript that an annotator can tell whether a rising pitch contour is the start of an H* accent or an LH% boundary. Pitch contour alone is not enough.

3. Syllable-level alignment and relative height are often important for drawing distinctions like H*/L+H*.

4. It is also important to know the context and the meaning of each utterance. Primary pitch accents invariably mark kontrast, the property of distinguishing the intended theme or rheme from a set of (established or implicated) alternatives. This is a property that the annotator can only decide by paying attention to the meaning of the text.

5. Many utterances consist of a single information unit, usually the rheme, but sometimes a marked theme. If there is a theme as well as a rheme present, the theme will usually be unmarked, and hence leave the pitch contour unaffected.

6. Since unmarked themes are characterized by low pitch without a detectable boundary, they leave thge position of the theme-rheme boundary (if any) ambiguous. A good test for whether a stretch of low pitch is a possible null theme is the following. *If it is possible in the context to utter it with the L+H* LHtheme tune, then its a possible theme*. If not, not.

3.2 APML

The APML DTD is as follows:

```
<!ELEMENT apml (turnallocation?, performative*, turnallocation?)>

<!ELEMENT turnallocation (performative)>
<!ATTLIST turnallocation type (take|give) #REQUIRED>

<!ELEMENT performative (#PCDATA|theme|rheme|emphasis|boundary|pause)*>
<!ATTLIST performative
type (greet|question|inform|paraphrase|suggest) "inform">

<!ELEMENT theme (#PCDATA|emphasis|boundary|pause)*>
<!ATTLIST theme affect (sorry-for|relief) #IMPLIED
                  belief-relation
(gen-spec|cause-effect|solutionhood|suggestion|modifier|justification) #IMPLIED>

<!ELEMENT rheme (#PCDATA|emphasis|boundary|pause)*>
<!ATTLIST rheme affect (sorry-for|relief) #IMPLIED
                  belief-relation
(gen-spec|cause-effect|solutionhood|suggestion|modifier|justification) #IMPLIED>

<!ELEMENT emphasis (#PCDATA)>
<!ATTLIST emphasis type (theme|rheme) #IMPLIED
                  level (strong|medium|weak) "medium"
                  x-pitchaccent (Hstar|Lstar|LplusHstar|LstarplusH|HstarplusL|HplusLstar) "Hstar"
                  deictic CDATA #IMPLIED
                  adjectival (small|tiny) #IMPLIED>
```

```
<!ELEMENT boundary EMPTY>
<!ATTLIST boundary type (L|H|LL|HH|LH|HL) "LL">

<!ELEMENT pause EMPTY>
<!ATTLIST pause sec CDATA #REQUIRED>
```

Of these data structures, those defining turn allocation and performative aspects do not affect speech synthesis, and are fairly self-explanatory. The reader is referred to de Carolis et al. (2004) for discussion. The present concern is with the assignment of the labels *theme*, *rheme*, *emphasis* and *boundary*, relating to information structure and intonation structure. (Certain further specifications of affect, propositional attitude, deixis and comparison which are relevant to gesture and facial expression are here ignored, again referring the reader to de Carolis et al. (2004) for discussion, since they do not affect speech synthesis.)

3.3 Some Examples

With the DTD as `apml.dtd` on the path or in the same directory, the APML examples given in the rest of the paper can be heard by making them into files prefixed by the following two lines:

```
<?xml version="1.0"?>
<!DOCTYPE apml SYSTEM "apml.dtd" []>
```

To do so, festival should be invoked, and the following two lines should be typed at the festival prompt.

```
(require 'apml)
(apml_initialise)
```

An APML file `myfile.apml` can then be played by typing the following incantation to the festival prompt:

```
(utt.play (apml_file_synth "myfile.apml"))
```

The easiest spoken texts to annotate are conversational diads consisting of a *wh*-question followed by its answer, of the kind used in (1) and (2). The answer (1) is annotated as follows:

(3)
```
<apml>
<performative type="inform">
<theme> He bit </theme>
<rheme><emphasis x-pitchaccent="Hstar">this</emphasis> finger <boundary type="LL"/>
</rheme>
</performative>
</apml>
```

Since the default for the emphasis pitch accent is H*, and the default boundary is LL%, a similar result can be obtained from the following variant.

(4)
```
<apml>
<performative type="inform">
<theme> He bit </theme>
<rheme><emphasis>this</emphasis> finger <boundary/>
</rheme>
</performative>
</apml>
```

Since Festival translates punctuation as boundaries, a similar result can be obtained by using a full stop instead of the boundary. This will be indistinguishable auditorily from the following "all rheme" version:

(5)
```
<apml>
<performative type="inform">
<rheme> He bit <emphasis>this</emphasis> finger.</rheme>
</performative>
</apml>
```

And in fact the rheme label can be omitted as well. However, if the emphasis is left unspecified, Festival will default to emphasis on the word "finger", with anomalous results.

(6)
```
<apml>
<performative type="inform">
He bit this finger.
</performative>
</apml>
```

The answer (2) is annotated as follows:

(7)
```
<apml>
<performative type="inform">
<theme><emphasis x-pitchaccent="LplusHstar">He</emphasis> bit <boundary type="LH"/></theme>
<rheme><emphasis x-pitchaccent="Hstar">this</emphasis> finger <boundary type="LL"/>
</rheme>
</performative>
</apml>
```

Using markup of even this simple kind correctly is crucial to applications and in particular dialog applications. The following example was generated as a syntactically impeccable translation of a response (in German) to the question "How about meeting on Monday afternoon?"

(8) I unfortunately have no time on Monday afternoon.

When fed to a standard text-to-speech system, this was realised with a default pitch accent on "afternoon", with distracting effects on the evaluators. This inappropriate intonation can be represented in APML as follows:

(9)
```
<apml>
<performative type="inform">
I unfortunately have no time on Monday <emphasis
x-pitchaccent="Hstar"> afternoon </emphasis><boundary type="LL"/>
</performative>
</apml>
```

What we want is the following[5]

(10)
```
<apml>
<performative type="inform">
<rheme>I unfortunately have no <emphasis x-pitchaccent="Hstar">time</emphasis> <boundary type="L"/>
<theme>on Monday afternoon<boundary type="LL"/>
</theme>
</performative>
</apml>
```

This is a good example to try out the null theme test on, by putting an L+H* LH% tune on " on MONDAY AFTERNOON", as follows:

(11)
```
<apml>
<performative type="inform">
<rheme>I unfortunately have no <emphasis
x-pitchaccent="Hstar">time</emphasis><boundary type="L"/></rheme>
<theme>on <emphasis x-pitchaccent="LplusHstar"> Monday </emphasis>
<emphasis x-pitchaccent="LplusHstar"> afternoon </emphasis> <boundary type="LH"/>
</theme>
</performative>
</apml>
```

The following is a more complicated constructed example of the same kind as (7), from Steedman (2000), to illustrate the way that alternative sets influence annotation:

(12) Q: I know that Marcel likes the man who wrote the MUSICAL.
But who does he ADMIRE?

A: (Marcel ADMIRES) (the woman who DIRECTED the musical)

$\underbrace{\qquad}_{\textit{Background}}\ \underbrace{\text{L}+\text{H}*\text{LH}\%}_{\textit{Kontrast}}\ \underbrace{\qquad}_{\textit{Background}}\ \underbrace{\text{H}*}_{\textit{Kontrast}}\ \underbrace{\text{LL}\%}_{\textit{Background}}$

$\underbrace{\qquad\qquad\qquad}_{\textit{Theme}}\qquad\underbrace{\qquad\qquad\qquad\qquad\qquad}_{\textit{Rheme}}$

The preceding discourse sets up a context with Marcel, the director, and a theme via the wh question. The response assigns an L+H* accent to *admires* marking the phrase as thematic and admires as in kontrast with another word in an implicated competing theme, followed by a LH% boundary, which marks the end of the phrase and means that the hearer "owns" or is committed to this theme. The lack of a primary pitch-accent on *Marcel* means that that word does not contribute to distinguishing this theme from the alternatives: all implicated themes are about Marcel.

The response also assigns an H* accent to the word *directed*, but not to *the musical*. The rheme alternative-set is thereby implicated to include only individuals related to the musical in various ways. The lack of a primary pitch-accent on *woman* does not implicate that all the individuals in the context are women, but rather that the property of directing the musical is enough to distinguish this individual from all alternative individuals. (That is, it entails that there

[5]It might appear that this problem could be finessed by using the heuristic strategy of automatically deaccenting recently mentioned words like "Monday" and "afternoon". However, recent mention does not always mean that a word is backgrounded. Moreover, the tactic will not work for such anaphoric alternatives to "on Monday afternoon" as "on that day" or "then".

is only one director of the musical.)

All of these implicatures are already established or consistent with the context, so the example succeeds. The corresponding APML is as follows:

(13)
```
<apml>
<performative type="inform">
<theme> Marcel
   <emphasis x-pitchaccent="LplusHstar">admires</emphasis><boundary type="LH"/>
</theme>
<rheme> the woman who <emphasis x-pitchaccent="Hstar">directed</emphasis>
 the musical<boundary type="LL"/>
</rheme>
</performative>
</apml>
```

Real texts are rarely so easy to handle. For a start, the participants normally have the context so much under control that explicitly marking kontrasting words in themes is unnecessary. Thus primary theme pitch accents are generally very reduced or entirely absent, as in (14a), or the entire theme is simply omitted, as in (14b).

(14) a. Marcel admires the woman who DIRECTED$_{H*}$ the musical.$_{LL\%}$
b. The woman who DIRECTED$_{H*}$ the musical.$_{LL\%}$

For this reason, marked themes are quite rare in real dialog, although they do occur in situations like airline ticket reservation, in which it is genuinely difficult for participants to be clear about which of a number of similar flights (themes) they are giving information about (rhemes).

It is important to realize that utterances are not required to contain both a theme and a rheme. They may consist solely of one or more rhemes, or one or more themes. In fact, the most common use of the marked theme tune L+H*LH is in "all-theme" utterances, like the following:

(15) There are seats in BUSINESS class
L+H* LH%

This utterance is to be contrasted with the all-rheme alternative:

(16) There are seats in BUSINESS class
H* LL%

Example (16) is a good answer to the question "where are there seats?", when it realises the following APML information structure:

(17)
```
<apml>
  <performative type="inform">
  <theme>there are seats</theme>
  <rheme>in
  <emphasis x-pitchaccent="Hstar"> business</emphasis> class
  <boundary type="LL"/></rheme>
  </performative>
</apml>
```

The rheme is "in business class", the speaker takes responsibility, and it is a full answer to the question.

However, (16) is quite inappropriate—in fact, rude—as a move in a dialog searching for an economy class seat, and where the client has a preference for saving money. In that context (15) is more appropriate, annotated as follows:

(18)
```
<apml>
  <performative type="inform">
  <theme>there are seats in
  <emphasis x-pitchaccent="LplusHstar"> business</emphasis> class
  <boundary type="LH"/></theme>
  </performative>
</apml>
```

It differs from (16) in making the information into a theme, linking it to the context, by contrasting business with economy, by saying that the hearer is responsible for it, and most importantly leaving the hearer to work out any rheme for themselves.

This is an "indirect speech act" use of the theme tune. The effect is much the same as a more "tentative" or "polite" version of the rhematic (16), and an annotator might wish to think of it in that way as a shorthand, and make extensive use of it. But it is important to know that there is a lot going on behind the scenes in utterances like this, just as there is in the verbal content. (Similar factors explain why "You will just have to travel business class" may not be a good sell, no matter how true.)

Longer examples of all-theme utterances may consist of several distinct intonational phrases, as in the following example concerning the possibility of obtaining a business-class seat for one of a particular set of flights:[6]

(19)
```
<apml>
  <performative type="instruct">
  <theme>there
    <emphasis x-pitchaccent="LplusHstar">are</emphasis>
    seats in business class <boundary type="LH"/>
  </theme>
  <theme>on the
    <emphasis x-pitchaccent="LplusHstar">British Airways</emphasis>
    flight<boundary type="LH"/>
  </theme>
 <theme> that arrives at
    <emphasis x-pitchaccent="LplusHstar">four</emphasis>
    <emphasis x-pitchaccent="LplusHstar">twenty</emphasis>
    <boundary type="LH"/> .
  </theme>
  </performative>
</apml>
```

An important point to remember in dealing with examples like this is that complex phrases are commonly broken up into several intonational phrases, and hence several information units, as in

[6]This example is from the Edinburgh FLIGHTS project corpus Moore et al. (2004), and many of the observations below arose in discussions with Mike White.

(20) (This is the cat)(that ate the rat)(that lived in the house)(that Jack built.)

This sort of prosody is appropriate when such PP and relative clause modifiers are *appositive* or attributive, rather than restrictive. Syntactically, appositives are sentential or VP level modifiers, rather than NP modifiers, and their link to the head noun ("seats," "flights," "cat," etc) is anaphoric, as can be inferred from the fact that such linking is immune to constraints such as subject island-hood:

(21) A man$_i$ came in who$_i$ was wearing a bowler hat.

Such appositive attributives are very common. Even if a unique British Airways flight has not been explicitly mentioned, and there are many flights operated by that company in the timetable, a speaker may utter (19) so long as only one of these flights is relevant to the needs of the hearer. So long as this is not inconsistent with their context, the hearer will effortlessly accomodate the existence of such a unique relevant British Airways flight.

Even if an annotator cannot see any boundaries instrumentally, they can be quite sure that to bracket example (19) as follows would be wrong:[7]

(22)
```
<apml>
  <performative type="instruct">
  <theme>there
    <emphasis x-pitchaccent="LplusHstar">are</emphasis> seats in
    business class on the
    <emphasis x-pitchaccent="LplusHstar">British Airways</emphasis>
    flight that arrives  <boundary type="LH"/>
  </theme>
  <rheme> at <emphasis x-pitchaccent="Hstar">four</emphasis>
   <emphasis x-pitchaccent="Hstar">twenty</emphasis> <boundary type="LH"/> .
  </rheme>
  </performative>
</apml>
```

The present section has so far been exclusively concerned with just two intonational tunes, because most of the intonational work in monologs and informative dialogs of the kind discussed above is done by H* accents (for rhemes) and L+H* accents (for themes). Rheme accents are most commonly associated with L or LL% boundaries, because a rheme is architypically the speakers way of moving the discourse forward, and it is natural for the speaker to be committed to and take responsibility for that contribution. Theme accents are most commonly combined with LH% boundaries, because the theme is architypically the link to the previous discourse, which is in normal circumstances shared by speaker and hearer. If it is necessary to mark contrast at all, then the contrast will normally be with an alternative that the hearer knows about, so it is natural for the speaker to mark it as the hearer's responsibility.

The latter observation can be exploited to indirect effect, and explains the prevalance of all-theme utterances like (15): by marking the contrast as the hearer's responsibility, despite the fact that the latter may never have thought of travelling business class until now, the hearer is encouraged to take the alternative on board.

[7] Steedman (2000) shows that such intonation is in fact forced by the fact that intonational phrases are sensitive to syntactic "island boundaries".

3.4 Other Prosodic Phrasal Tunes

The following section concerns some more rarified and and discourse-pragmatically specialized tunes that will rarely be called for in practical applications, but which will occasionally be found in corpora of real dialog.[8]

Under certain circumstances, the natural tendency of rhematic H* to associate with speaker-oriented LL% and for thematic L+H* to associate with hearer-oriented LH% may be overridden. An LH% boundary on an H*-marked rheme is appropriate for "echo questions," as in the following response to A's statement that she has just bought a painting by Mondrian, because the rheme is indeed still the hearer's responsibility.

(23) B: A MONDRIAN ?
H* LH%

In fact, a speaker-responsible LL% would be anomalous in this situation, since such an utterace would not advance the conversation at all as far as the hearer goes. It is only by the speaker *not* taking the responsibility that the utterance has the effect of requesting confirmation of a surprising rheme. This contour could usefully be deployed in generating echo questions to check the results of imperfect speech recognition for interactive systems.

Howevcer, because it marks the rheme as something the hearer is already committed to, the H*LH% tune can convey impatience, to the point of rudeness. Compare the following responses to A's question "Where are my socks?"

(24) B: They're in your DRAWER .
H* LL%

(25) B: They're in your DRAWER !
H* LH%

Such a contour may also be used indirectly, as in utterances like the following:

(26) B: I'm WARNING you !
H* LH%

Such an utterance differs from the literal statement that would result from using an LL% boundary. By marking this rheme (counterfactually) as already the hearer's responsibility, the utterance actually becomes an act of warning or threatening.

Clearly, the H*LH% contour should be deployed with caution in standard dialogue applications such as airline flight booking, and be strictly confined to echo questions.

The other major rheme accent, L*, differs from H* in marking the rheme as not agreed. It is therefore natural for it to occur with the LH% boundary that marks hearer responsibility. The following response to A's earlier statement about buying a Mondrian expresses incredulity, as well as requesting confirmation:

(27) B: A MONDRIAN ?
L* LH%

Substituting the H+L* pitch accent for L* (that is, starting with high pitch) and/or substituting

[8] Some tones discussed below are not currently deployed in Festival, so we omit APML markup in this section.

the HH% boundary for LH% increase the effect of incredulity by emphasising the degree of non-agreement and hearer responsibility via the pitch-range.

(Exagerating the pitch range of H* has a similarly agreement-emphasising effect, and is often used in pedagogical discourse, as Pierrehumbert and Hirschberg (1990) point out using the following example:[9]

(28) A: The hint key gives you hints if you need help.
A: HIT the HINT key !
H* H* LL%

The use of the L*LH% contour is common in questions, particularly where these constitute indirect speech acts such as polite requests,

(29) A: Can you tell me the TIME ?
L* LH%

By marking the ability precondition as not agreed, and as the hearers responsibility, the speaker avoids implying that the hearer will necessarily comply. This usage can probably be safely deployed in dialog applications involving questions and requests, but it is not as simple as it looks.

The remaining accent to discuss is the other theme pitch accent, L*+H. Like all L* accents, it marks non agreement. Since non-agreed themes are incompatible with progressing the discourse via a rheme, as normal, L*+H LH% theme tunes are rarely if ever used in standard theme-rheme utterances like (2). However, they are commonly used in all-theme utterances such as (15), which then appears as follows:

(30) There are seats in BUSINESS class!
L*+H LH%

The effect of marking non-agreement is appropriate for indirectly denying a claim such as that there are no seats available anywhere. Substituting a speaker-responsible LL% (or HL%) makes the disagreement final:

(31) There are seats in BUSINESS class!
L*+H LL%

A very common all-theme utterance with this pitch accent, indicating non agreement on the speakers ability to answer a question, is the following:

(32) *I* d'know !
L*+H LH%

The use of such agressive tunes in standard dialogue applications is probably inadvisable.

3.5 A Note on the Boundary Tones

I have said very little about the specific effect of alternative boundaries above, except to divide them into the speaker- and hearer responsible classes, because the theory is underdeveloped on this point, and my recommendations here are correspondingly tentative. The main

[9] Pierrehumbert and Hirschberg give these accents as H*+L. The notation is changed to fit the more recent ToBI conventions.

difference seems to lie in retention or yielding of the turn. Of the boundaries marking speaker-responsibility, L and LL% seem to be on a continuum, differing only in the degree of lengthening they induce, and appear to retain the turn for the speaker. (That is not of course to say that the hearer may not claim the turn anyway.) On the other hand, the boundary HL%, which "trails off" at the middle level, while still taking speaker responsibility, seems to explicitly relinquish the turn. It is therefore naturally used in summoning or "calling" with the expectation of a response as in the following example from Ladd (who also points out that this expectation makes this tune incompatible with warnings such as "Fire!"):[10]

(33) Jacob . Your LUNCH is ready .
H* HL% H* HL%

Pierrehumbert and Hirschberg (1990) give the following examples of HL%, to convey an obvious but forgotten fact explaining a baby's wakefulness, and news of an unsurprising but mildly unwelcome event:

(34) SHE'S TEETHING .
H+L* HL%

(35) Oh DARN it .
H+L* HL%

Among the boundaries that mark hearer-commitment according to the present theory, H and LH% do not seem to explicitly yield the turn, although marking something as the hearer's responsibility often has the effect of calling for a response. However, the HH% boundary explicitly yields the turn, and is therefore extremely common in yes-no questions and certain varieties of *wh*-question like the following, where the speaker indicates thereby that non-response is not an option:

(36) WHAT did you say ?
H+L* HH%

4 Using APML to Mark Up Content for Meaning-to-Speech Synthesis

All of the rules of thumb that were applied above to corpus annotation should be kept in mind when annotating text for synthesis or generating such texts from scratch. In addition, three more points are worth remembering.

1. Less is more. If in doubt about the correct analysis, do as little as possible. APML is designed to default to something statistically likely at every level so you may be lucky.

2. In particular, don't insist on putting a marked theme in every utterance. Sentences with both a rheme and a marked theme are quite rare. And unmarked theme boundaries won't affect the synthesis. If in doubt, make it an all-rheme utterance.

3. Where you do decide to intervene prosodically, the most important thing to do is to position the primary pitch accents correctly. This depends exclusively on the structure of the

[10] Again Pierrehumbert and Hirschberg's notation for this example is different, using H*+L accents under a pre-ToBI convention.

alternatives sets. If you or your program has a good model of the discourse context, this is relatively straightforward. Getting a pitch accent on the wrong word, say by relying on defaults too much, is much more disastrous than assigning the wrong pitch accent contour.

4. Boundaries should be thought of as solely affecting pitch and/or vowel length. If you want a break or pause, either use puctuation after the boundary (comma is shorter than stop etc.), or use a `<pause>` element to control breaks with explicit durations.

5. If the speech that Festival produces from your mark-up is not what you expected, the cause may well be the lexicon. Festival uses a pronunciation model that is not always correct for unknown words. For example, the apml (13) for example (12) may sound better if you fix the lexical entry for "admires". Words like "fire," "admire," and "film" effectively have two syllables-worth of "-ire" and "-ilm", as in "fiya" and "fillum." Looking at the lexical entry for "admires" shows that Festival thinks it has two syllables:

   ```
   festival> (lex.lookup 'admires)
   ("admires" nil (((ae d) 0) ((m ay r z) 1)))
   ```

   This can be fixed as follows, either interactively or in a file, and may improve things.

   ```
   festival> (lex.add.entry '("admires" nil (((ae d) 0) ((m ay) 1)(( ax z) 0))))
   nil
   ```

The following is an example of how to mark up by hand a semiautomatically generated text from the MagiCster project, adhering to these simple maxims, illustrating many of the uses of intonation that a good intonational content generator should strive for. The speaker is a doctor who is reporting the results of a test to a patient.

(37) Good morning, Mr Smith. I'm sorry to tell you that you have been diagnosed as suffering from a mild form of what we call angina pectoris. This is a spasm of the chest, resulting from overexertion when the heart is diseased. To solve this problem, there are two drugs I would like you to take. The first one is Aspirin, which is an analgesic. That is, it relieves the pain. I have prescribed it to cure your angina. The only problem is that this drug can be associated with some sideeffects. Right?

The first sentence, like many discourse-initial sentences, is an all-rheme utterance. The fact that the hearers name is Smith is background, so the utterance annotated as follows:

(38)
```
<performative type="greet">
<rheme>Good <emphasis x-pitchaccent="Hstar">morning </emphasis>
       Mr Smith<boundary type="LL"/>
</rheme>
</performative>
```

The second sentence consists of a theme, or link to the previous discourse, followed by a rheme, giving some news. The theme is essentially that the news is bad, and it contrasts with the possibilities of conveying good news, or not conveying the news, so "sorry" and "tell" get theme accents. (It is possible to only accent one or the other, when subtly different theme-alternative

sets are conveyed.) The rest of the sentence is a rheme. As is common with didactically informative rhemes like this one, virtually every content word gets a primary rheme accent. The only possiblity for dropping emphases seems to be on the word "diagnosed" (since diagnosis might be held to be background.

(39)
```
<performative type="inform">
<theme>
I'm sorry to <emphasis x-pitchaccent="LplusHstar">tell</emphasis> you
<boundary type="LH"/>
</theme>
<rheme>that you have been <emphasis x-pitchaccent="Hstar"> diagnosed </emphasis>
       as <emphasis x-pitchaccent="Hstar">suffering</emphasis>
       from a <emphasis x-pitchaccent="Hstar"
                        adjectival="small">mild</emphasis>
       <emphasis x-pitchaccent="Hstar">form</emphasis>
       of what we call <emphasis x-pitchaccent="Hstar">angina pectoris</emphasis>.
       <boundary type="LL"/>
</rheme>
</performative>
```

The third sentence elaborates the meaning of the term "angina pectoris". It consists of two theme-rheme pairs, separated by a comma. The first theme links the utterance to the previous one via the anaphor "this". The latter word could attract an L+H* theme accent, but the context is unambiguous enough to make such an accent unnecessary. The following rheme again gets an H* accent on every content word. (Notice that while commas are often associated with LH%, which is often referred to as a "continuation rise," in this case this would result in marking the rheme as hearer-committed, which is generally inconsistent with didactic rhemes of this kind.) The second theme links the second rheme as a further attribute of angina pectoris, unmarked because it is background that symptoms have causes. Again, all content word in the rheme are accented except "diseased," presumably since disease is background when we talk about hearts in this context.

(40)
```
<performative type="inform">
<theme>
       This is
</theme>
<rheme>
        a <emphasis x-pitchaccent="Hstar"> spasm </emphasis>
        of the <emphasis x-pitchaccent="Hstar"> chest </emphasis>
       <boundary type="LL"/>,
</rheme>
<theme>
       resulting from
</theme>
<rheme>
       <emphasis x-pitchaccent="Hstar">overexertion</emphasis>  when the <emphasis x-pitchaccent="H
       is diseased
       <boundary type="LL"/>.
</rheme>
</performative>
```

The fourth sentence consists of a marked theme, in which "solve" is the only word to get a theme accent (since it is background that angina is a problem) and the implicated alternative is *not* solving the problem, and a rheme in which the words that get an accent are those which distinguish the treatment from others, while the modifier "(that) I would like you take" is deaccented (since wanting you to take them is the usual reason for doctors telling you about drugs.)

(41)
```
<performative type="inform">
<theme>
       To <emphasis x-pitchaccent="LplusHstar">solve </emphasis>
       this problem
       <boundary type="LH"/>
</theme>
<rheme>
       there are <emphasis x-pitchaccent="Hstar">two </emphasis>
       <emphasis x-pitchaccent="Hstar">drugs </emphasis>I would like you to take
       <boundary type="LL"/>
</rheme>
</performative>
```

The succeeding utterances (of which we will only discuss the first) discuss the virtues and disadvantages of the two drugs in turn. This is the classic use for a marked theme, as was seen in connection with the early example (2), so the utterance begins wqith a marked theme in which the word "first", which distinguishes this theme from that of the second of the two drugs,gets the accent. The first rheme is so simple as to call for no comment. However, it is followed by a second theme rheme pair, forming an attributive relative, or anaphoric modifier of "aspirin". In such attributive relatives (which generally call for a comma's worth of pause, unlike restrictive relatives), prefixes like "which is," "which the textbooks say is," and the like, are themes, linking backwards, in this case unmarked. Again the attributive rheme is straightforward..

(42)
```
<performative type="inform">
<theme> The <emphasis x-pitchaccent="LplusHstar">first</emphasis>
        one
        <boundary type="LH"/>
</theme>
<rheme>is <emphasis>Aspirin</emphasis> <boundary type="LL"/></rheme>,


<theme belief-relation="modifier">
       which is
</theme>
<rheme>an <emphasis>analgesic</emphasis><boundary type="LL"/></rheme>
</performative>
```

The next utterance is also elaboration of the attributes of to aspirin. The phrase, “that is” is linking, hence thematic. The rheme is as usual.

(43)
```
<performative type="inform">
<theme>
       That is,
</theme>
<rheme>it <emphasis x-pitchaccent="Hstar">relieves</emphasis>
       the <emphasis x-pitchaccent="Hstar">pain.</emphasis>
       <boundary type="LL"/>.
</rheme>
</performative>
```

The next utterance also states a property of aspirin. However, this time it is a negative problem, so a marked theme is called for in which the word “problem” distinguishes the theme, and gets an accent. (The word “only” can also be contrasted, to implicate an alternative theme concerning multiple problems.) In the rheme, the idea that the drug is associated with the property in question is background: otherwise content words get a pitch accent as usual.

(44)
```
<performative type="inform">
<theme> The only
            <emphasis x-pitchaccent="LplusHstar">problem
            </emphasis> is
        <boundary type="LH"/>
</theme>
<rheme>that this drug can be associated with
some <emphasis x-pitchaccent="Hstar">sideeffects</emphasis>
       <boundary type="LL"/>
</rheme>
</performative>
```

At this point, the speaker yields the turn, before proceeding to discuss the second drug, in order to invite questions.

(45)
```
<performative type="question">
<rheme><emphasis x-pitchaccent="Lstar">Right</emphasis> <boundary type="LH"/>?</rheme>
</performative>
```

The following is the same text in one piece, marked up for intonation using some of the alternatives mentioned above, and including some other elements of APML markup relevant to facial and gestural expression discussed in de Carolis et al. 2004.

```
<apml>

<turnallocation type="take">
<performative type="greet">
<rheme>Good <emphasis x-pitchaccent="Hstar">morning </emphasis>
       Mr Smith<boundary type="LL"/>
</rheme>
</performative>
</turnallocation>


<performative type="inform">
<theme belief-relation="gen-spec"
               affect="sorry-for">
       I'm <emphasis x-pitchaccent="LplusHstar">sorry</emphasis> to <emphasis x-pitchaccent="LplusHstar"
<boundary type="LH"/>
</theme>
<rheme>that you have been <emphasis x-pitchaccent="Hstar"> diagnosed </emphasis>
       as <emphasis x-pitchaccent="Hstar">suffering</emphasis>
       from a <emphasis x-pitchaccent="Hstar"
                        adjectival="small">mild</emphasis>
       <emphasis x-pitchaccent="Hstar">form</emphasis>
       of what we call <emphasis x-pitchaccent="Hstar">angina pectoris</emphasis>.
       <boundary type="LL"/>
</rheme>
</performative>

<performative type="inform">
<theme belief-relation="gen-spec">
       This is
</theme>
<rheme>
        a <emphasis x-pitchaccent="Hstar"> spasm </emphasis>
        of the <emphasis x-pitchaccent="Hstar"
                         deictic="0 30 0"> chest </emphasis>
       <boundary type="LL"/>,
</rheme>
<theme belief-relation="gen-spec">
       resulting from
</theme>
<rheme belief-relation="cause-effect">
       overexertion when the <emphasis x-pitchaccent="Hstar"> heart </emphasis>
       is diseased
       <boundary type="LL"/>
</rheme>
```

```
</performative>

<performative type="suggest">
<theme belief-relation="solutionhood">
       To <emphasis x-pitchaccent="LplusHstar">solve </emphasis>
       this problem
       <boundary type="LH"/>,
</theme>
<rheme belief-relation="suggestion">
       there are two
       <emphasis x-pitchaccent="Hstar">drugs </emphasis>I would like you to take.
       <boundary type="LL"/>
</rheme>
</performative>


<performative type="inform">
<theme> The <emphasis x-pitchaccent="LplusHstar">first</emphasis>
        one
        <boundary type="LH"/>
</theme>
<rheme>is <emphasis>Aspirin,</emphasis> <boundary type="LL"/></rheme>


<theme belief-relation="modifier">
       which is
</theme>
<rheme>an <emphasis>analgesic.</emphasis><boundary type="LL"/></rheme>
</performative>

<performative type="paraphrase">
<theme belief-relation="gen-spec"
              affect="relief">
       that is
</theme>
<rheme>it <emphasis x-pitchaccent="Hstar">relieves</emphasis>
       the <emphasis x-pitchaccent="Hstar">pain.</emphasis>
       <boundary type="LL"/>
</rheme>
</performative>

<performative type="inform">
<theme> I have <emphasis x-pitchaccent="LplusHstar">prescribed</emphasis> it
        <boundary type="LH"/>
</theme>
<rheme belief-relation="justification">
       to cure your <emphasis x-pitchaccent="Hstar">angina </emphasis>
       <boundary type="LL"/>
```

```
</rheme>
</performative>

<performative type="inform">
<theme affect="sorry-for"> The <emphasis x-pitchaccent="LplusHstar"
                      adjectival="tiny">only
    </emphasis>
            <emphasis x-pitchaccent="LplusHstar"
                      adjectival="tiny">problem
            </emphasis>
         <boundary type="LH"/>
</theme>
<rheme>is that this drug can be associated with
some <emphasis x-pitchaccent="Hstar">sideeffects</emphasis>
       <boundary type="LL"/>.
</rheme>
</performative>

<turnallocation type="give">
<performative type="question">
<rheme><emphasis x-pitchaccent="Lstar">Right</emphasis> <boundary type="LH"/>?</rheme>
</performative>
</turnallocation>
</apml>
```

Finally, if you are so pleased with the results of the synthesis that you want to make a sound file of the result, then the following instructions may be useful:

```
festival> (set! foo  (utt.play (apml_file_synth "filename.apml")))
#<Utterance 0x40a538b8>
festival> (utt.save.wave foo "filename.wav" 'riff)
#<Utterance 0x40a538b8>
festival>
```

Further facilities are described in the Festival documentatiion.

Acknowledgements

I am grateful to Sasha Calhoun, Rob Clark, Steve Isard, Bob Ladd, Johanna Moore, and Mike White for many discussions. Sasha and Mike road-tested many of the annotation instructions and claims, and drew many pitfalls to my attention. The work was supported in part by EU FET grant MagiCster, EPSRC grant GR/R02450, and the Edinburgh-Stanford Link grant "Sounds of Discourse" from Scottish Enterprise.

References

Bolinger, Dwight. 1958. "A Theory of Pitch Accent in English." *Word*, 14, 109–149. Reprinted in Bolinger (1965), 17-56.

Bolinger, Dwight. 1961. "Contrastive Accent and Contrastive Stress." *Language*, 37, 83–96. Reprinted in Bolinger (1965), 101-117.

Bolinger, Dwight. 1965. *Forms of English.* Cambridge, MA: Harvard University Press.

Büring, Daniel. 1997a. "The Great Scope Inversion Conspiracy." *Linguistics and Philosophy*, 20, 175–194.

Büring, Daniel. 1997b. *The Meaning of Topic and Focus: The 59th Street Bridge Accent*. London: Routledge.

Calhoun, Sasha. 2004. "Phonetic Dimensions of Intonational Categories." In *Proceedings of Prosody 2004, Nara, Japan, March.* Poster: to appear. cf. `http://www.cstr.ed.ac.uk/~s0199920/research.html`.

de Carolis, Bernadina, Catherine Pelachaud, Isabella Poggi, and Mark Steedman. 2004. "APML, a Mark-up Language for Believable Behavior Generation." In H. Prendinger, ed., *Life-like Characters. Tools, Affective Functions and Applications*, 65–85. Berlin: Springer.

Cassell, Justine, Catherine Pelachaud, Norm Badler, Mark Steedman, Brett Achorn, Tripp Becket, Brett Douville, Scott Prevost, and Matthew Stone. 1994. "Animated Conversation: Rule-based Generation of Facial Expression, Gesture and Spoken Intonation for Multiple Conversational Agents." In *Proceedings of the ACM SIGGRAPH '94 Conference, Orlando FL*, 413–420.

Chafe, Wallace. 1974. "Language and consciousness." *Language*, 50, 111–133.

Clark, Herbert. 1996. *Using Language*. Cambridge: Cambridge University Press.

Clark, Herbert, and Catherine Marshall. 1981. "Definite Reference and Mutual Knowledge." In Aravind Joshi, Bonnie Webber, and Ivan Sag, eds., *Elements of Discourse Understanding*, 10–63. Cambridge: Cambridge University Press.

Cohen, Philip. 1978. *On Knowing What to Say: Planning Speech Acts*. Ph.D. thesis, University of Toronto.

Gundel, Janet. 1974. *The Role of Topic and Comment in Linguistic Theory*. Ph.D. thesis, University of Texas, Austin.

Gundel, Janet, and Torsten Fretheim. 2001. "Topic and Focus." In Laurence Horn and Gregory Ward, eds., *Handbook of Pragmatic Theory*. Oxford: Blackwell.

Gunlogson, Christine. 2001. *True to Form: Rising and Falling Declaratives in English.* Ph.D. thesis, University of California at Santa Cruz. In preparation.

Gunlogson, Christine. 2002. "Declarative Questions." In Brendan Jackson, ed., *Proceedings of* Semantics and Linguistics Theory XII, 144–163. Ithaca, NY: Cornell University. To appear.

Gussenhoven, Carlos. 1983. *On the Grammar and Semantics of Sentence Accent*. Dordrecht: Foris.

Hajičová, Eva, and Petr Sgall. 1988. "Topic and Focus of a Sentence and the Patterning of a Text." In J änos Petöfi , ed., *Text and Discourse Constitution*, 70–96. Berlin: de Gruyter.

Halliday, Michael. 1963. "The Tones of English." *Archivum Linguisticum*, 15, 1.

Halliday, Michael. 1967a. *Intonation and Grammar in British English*. The Hague: Mouton.

Halliday, Michael. 1967b. "Notes on Transitivity and Theme in English, Part II." *Journal of Linguistics*, 3, 199–244.

Hobbs, Jerry. 1990. "The Pierrehumbert-Hirschberg Theory of Intonational Meaning Made Simple: Comments on Pierrehumbert and Hirschberg." In Philip Cohen, Jerry Morgan, and Martha Pollack, eds., *Intentions in Communication*, 313–323. Cambridge, MA: MIT Press.

Karttunen, Lauri, and Stanley Peters. 1979. "Conventional Implicature." In Choon-Kyu Oh and David Dinneen, eds., *Syntax and Semantics 11: Presupposition*, 1–56. New York: Academic Press.

Kartunnen, Lauri. 1976. "Discourse Referents." In J. McCawley, ed., *Syntax and Semantics*, vol. 7, 363–385. Academic Press.

Kruijff-Korbayová, Ivana, and Mark Steedman. 2001. "Discourse and Information Structure." *Journal of Logic, Language, and Information*, 12, 249–259.

Ladd, D. Robert. 1980. *The Structure of Intonational Meaning*. Bloomington: Indiana University Press.

Ladd, D. Robert. 1996. *Intonational Phonology*. Cambridge: Cambridge University Press.

Ladd, D. Robert, and Astrid Schepman. 2003. "'Sagging transitions' between high pitch accents in English: Experimental evidence." *Journal of Phonetics*, 31, 81–112.

Lewis, David. 1969. *Convention: a Philosophical Study*. Cambridge MA: Harvard University Press.

Lewis, David. 1979. "Scorekeeping in a Language Game." *Journal of Philosophical Logic*, 8, 339–359.

Liberman, Mark. 1975. *The Intonational System of English*. Ph.D. thesis, MIT. Published by Garland Press, New York, 1979.

Moore, Johanna, Mary Ellen Foster, Oliver Lemon, and Michael White. 2004. "Generating Tailored, Comparitive Descriptions in Spoken Dialogue." In *Proceedings of the 17th International FLAIRS Conference(FLAIRS-04)*. AAAI.

Pierrehumbert, Janet, and Julia Hirschberg. 1990. "The Meaning of Intonational Contours in the Interpretation of Discourse." In Philip Cohen, Jerry Morgan, and Martha Pollack, eds., *Intentions in Communication*, 271–312. Cambridge, MA: MIT Press.

Pitrelli, John, Mary Beckman, and Julia Hirschberg. 1994. "Evaluation of Prosodic Transcription Labeling Reliability in the ToBI Framework." In *Proceedings of the 1st International Conference on Spoken Language Processing, Yokohama, Japan*, 123–126.

Prevost, Scott. 1995. *A Semantics of Contrast and Information Structure for Specifying Intonation in Spoken Language Generation*. Ph.D. thesis, University of Pennsylvania.

Prevost, Scott, and Mark Steedman. 1994. "Specifying Intonation from Context for Speech Synthesis." *Speech Communication*, 15, 139–153.

Rooth, Mats. 1985. *Association with Focus*. Ph.D. thesis, University of Massachusetts, Amherst.

Rooth, Mats. 1992. "A Theory of Focus Interpretation." *Natural Language Semantics*, 1, 75–116.

Selkirk, Elisabeth. 1984. *Phonology and Syntax*. Cambridge, MA: MIT Press.

Silverman, Kim, Mary Beckman, John Pitrelli, Marie Ostendorf, Colin Wightman, Patti Price, Janet Pierrehumbert, and Julia Hirschberg. 1992. "ToBI: A Standard for Labeling English Prosody." In *Proceedings of the International Conference on Spoken Language Processing, Banff, Alberta*, 867–870. Edmonton: University of Alberta.

Steedman, Mark. 2000. "Information Structure and the Syntax-Phonology Interface." *Linguistic Inquiry*, 34, 649–689.

Steedman, Mark. 2004. "Information-Structural Semantics for English Intonation." In Matt Gordon, Daniel Büring, and Chungmin Lee, eds., *LSA Summer Institute Workshop on Topic and Focus, Santa Barbara July 2001*. Dordrecht: Kluwer. To appear.

Stone, Matthew. 1998. *Modality in Dialogue: Planning Pragmatics and Computation*. Ph.D. thesis, University of Pennsylvania.

Taylor, Paul. 2000. "Analysis and Synthesis of Intonation Using the Tilt Model." *Journal of the Acoustical Society of America*, 107, 1697–1714.

Thomason, Richmond. 1990. "Accomodation, Meaning, and Implicature." In Philip Cohen, Jerry Morgan, and Martha Pollack, eds., *Intentions in Communication*, 325–363. Cambridge, MA: MIT Press.

Vallduví, Enric. 1990. *The Information Component*. Ph.D. thesis, University of Pennsylvania.

Vallduví, Enric, and Maria Vilkuna. 1998. "On Rheme and Kontrast." In Peter Culicover and Louise McNally, eds., *Syntax and Semantics, Vol. 29: The Limits of Syntax*, 79–108. San Diego, CA: Academic Press.