

3

Intonation in British English

DANIEL HIRST

1. Background

English is the official language of nearly 50 different countries and is currently spoken as a first language by over 300 million people (Crystal 1988). Among the numerous dialects of English spoken throughout the world, two, usually referred to as (Standard) American English and (Standard) British English, have a rather special status in that they are considered distinct standards for the teaching of English as a foreign language. Both dialects of English are spoken with a number of different accents.

For British English, one particular accent: “Received Pronunciation”, or “RP” (for a detailed description see Gimson 1962), traditionally defined as the accent of those educated in public schools, is generally presented as a model for foreign learners as well as a standard for BBC newsreaders. It has been estimated that the proportion of the population of England who actually speak RP is as small as 3% (Hughes and Trudgill 1979). It has been suggested (Brown 1977 p. 12) that RP today should be given a wider interpretation to include all speakers of “educated Southern English”. It does seem fairly safe to assume that the intonation system of RP is common to a rather wider section of the native population of (particularly Southern) Britain and it is to this system to which I shall refer below (unless otherwise stated) as “British English Intonation” but considerably more research into the intonation of other accents (see §3.1 below)

will obviously be necessary before we shall be in a position to claim, as Palmer (1922) did, that we are describing:

that system of intonation which is used by most of the natives of England. (p. ix)

1.1 General prosodic characteristics

There is a considerable literature on the nature of word stress in English and its relation to the segmental structure of the word (cf. Kingdon 1958a, Chomsky and Halle 1968, Guierre 1979, Hayes 1984, Fudge 1984, Halle and Vergnaud 1986). In particular, in contrast with earlier work which held that the position of word stress is entirely unpredictable, it has been argued in the framework of generative phonology that English word stress can be accounted for by a restricted number of fairly general phonological rules with lexical idiosyncrasies being reduced essentially to marking the final syllables of certain words as either inherently stressed or as **extrametrical** (*i.e.* invisible for the stress rules).

Typologically, English has a hybrid stress-system: on the level of the word, stress rules are in many ways similar to those of Romance languages in that the pattern of stress is basically determined with reference to the right edge of the word (with stress on the penultimate or antepenultimate syllable); Germanic suffixes, however, such as *-ing* and *-ly*, generally do not affect stress, and compound words in English, as in other Germanic languages, are generally stressed on the initial element.

Some authors distinguish more than two degrees of stress/accent. The final or most prominent pitch accent of an intonation unit is often referred to as carrying **primary stress**; a syllable which contains a full rather than a reduced vowel is sometimes said to carry **tertiary stress**. Whether or not an accented syllable is manifested by pitch or solely by duration and/or loudness is sometimes treated as a further degree of accentuation. For this chapter, I assume (following Bolinger 1958) one binary distinction between accented and unaccented syllables on the level of phonetic realisation and another binary distinction between stressed and unstressed syllables on the level of lexical representation.

1.2 Outline of the approach adopted in the chapter

In the following section I shall not present any new data on the intonation of British English but attempt simply to give a brief guide to what seem to me some significant results from the vast and ever-growing literature on the subject. The most exhaustive description of British English intonation is that of Crystal (1969). For discussion of more recent work see Couper-Kuhlen 1986, Cruttenden 1986. In §3.1, I outline some important differences which have been described in the intonation systems of other accents of the British Isles and in §4

I attempt to show how these descriptions can relate to a more general phonological theory of intonation.

2. Description of intonation patterns

2.1 Description of a basic non-emphatic pattern

a. Rhythmic structure

A number of linguists have claimed that intonation patterns are best described by means of a hierarchically organised structure with syllables being grouped into higher order prosodic constituents, each containing one accented syllable. This is often referred to as the **foot**, following Abercrombie (1964), who, in an extremely influential article, borrowed the term from traditional poetics to describe a sequence of syllables containing one stressed syllable followed by any number of unstressed syllables. Abercrombie specifically claimed that the foot is “independent of word boundaries” (p. 17) so that a sentence like

(1) a They preDICTed his eLEction.

would be analysed as:

(1) b They pre- | DICted his e- | LEction. |

where "|" corresponds to foot boundaries.

A more sophisticated model of rhythmic structure had earlier been proposed by Jassem (1952), according to which English speech is organised into two kinds of units: the **Narrow Rhythm Unit**, which like the Abercrombian foot consists of a stressed syllable followed by a sequence of unstressed syllables, and the **Anacrusis** consisting of a sequence of proclitic unstressed syllables. Anacrusis and Narrow Rhythm Unit combine to constitute the **Total Rhythm Unit**. Jassem claimed in particular that the rhythmic organisation of these two types of constituents is completely different: unstressed syllables in the Anacrusis tend to be pronounced “extremely rapidly” whereas the duration of each syllable in a Narrow Rhythm Unit tends to be inversely proportional to the number of syllables in that unit giving rise to the impression of isochrony which has often been attributed to languages like English (Lehiste 1977, Adams 1979). In a statistical analysis of a corpus containing both isolated sentences and a continuous dialogue (Units 30 and 39 of Halliday 1970), Jassem *et al.* (1984) present persuasive evidence that Jassem’s model gives a better account of durational patterns of English rhythm than does the Abercrombian model. Jassem’s model also appears compatible with data from a recent corpus study based on a twenty-minute continuous recording of a short story read by a professional actor (Campbell 1992 and personal communication). It is not, however, clear whether it is possible to account for both rhythm and melody by a single model of prosodic structure. In particular, it is noteworthy that Jassem

(1952) makes no use of Total Rhythm Units in his description of the melodic patterns of English but instead groups Anacrusis with the preceding Narrow Rhythm Unit (cf. pp. 49–50) to form **Tonal Units** which are said to be the domain of accentual pitch movements. The prosodic structure implied by this model can consequently be summarised as follows:

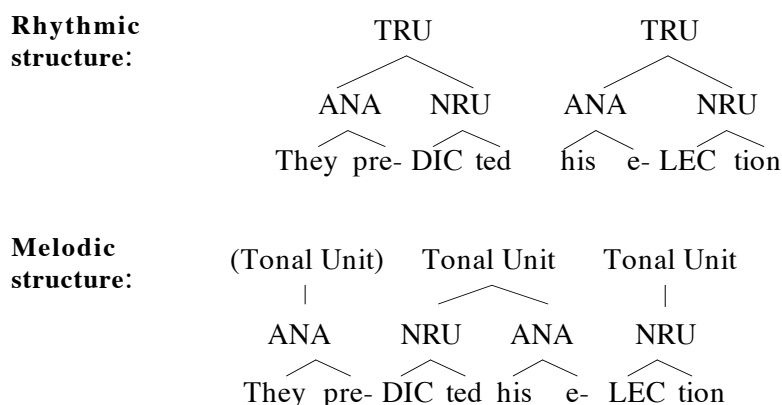


Figure 1. Jassem's model of rhythmic and melodic structure: TRU = Total Rhythm Unit; NRU = Narrow Rhythm Unit; ANA = Anacrusis.

This is obviously a field in which much work still remains to be done, in particular in so far as cross-language studies are concerned. In the rest of this chapter (cf. also Hirst and Di Cristo 1984; Hirst 1987) I shall use Jassem's term **Tonal Unit** to refer to an appropriate constituent for modelling intonation patterns on this level with the proviso that a more complex structure such as that discussed above may eventually prove necessary for a more complete description.

Above the level of the Tonal Unit, a further level of prosodic structure is generally considered necessary, referred to variously as the **Tone Group** (Palmer 1922; Schubiger 1958; Halliday 1967a, 1970; Gussenhoven 1984) the **Tune** (Armstrong and Ward 1926; Schubiger 1935; Jassem 1952; Kingdon 1958) the **Tone Unit** (Crystal 1969; Couper-Kuhlen 1986) or the **Intonation Group** (Cruttenden 1986). In parallel with the term **Tonal Unit** as defined above I shall refer to this higher-level structure here as the **Intonation Unit** (Hirst and Di Cristo 1984; Hirst 1987).

b. Stress and accent

The final accent of an Intonation Unit has often been given special status in descriptions and has been referred to since Palmer (1922) as the **nucleus**. For a historical account of the concept of nucleus in intonation studies see Cruttenden

(1990). The nucleus can occur before the last stressable syllable without in any way implying contrastiveness or emphasis. Thus in a dialogue like:

- (2) a – I’ve got some nice lamb chops for lunch.
- b – I’m sorry, I’m afraid I don’t eat meat.

there is no contrasting sentence “I (VERB) meat” which (2b) is intended to imply: the final accent falls on *eat* by default, simply because *meat* is de-accented. For discussion of “default accent” cf. Ladd (1980), Fuchs (1984).

Another very interesting case consists of sentences with intransitive verbs which can be pronounced with either a single accent on the subject or with one accent on the subject and another on the verb (cf. Schmerling 1976, Ladd 1980, 1983) Allerton and Cruttenden (1979), Gussenhoven (1984), Faber (1987). All these studies show that accenting and de-accenting imply a complex linguistic process which, despite considerable research, is still not fully understood but which cannot be reduced to a simple question of informativeness or predictability. For a contrary point of view see Bolinger (1989).

c. Tonal structure

One of the simplest accounts of British English intonation was that of Armstrong and Ward (1926) who proposed to analyse the intonation of unemphatic sentences by means of two “tunes”. The first of these is said to be used in “ordinary, definite, decided statements (word, phrase or sentence)” (p. 9) while the second is said to be used mainly with questions, requests and incomplete groups (p. 22). Tune 2 will be dealt with in §2.2 below; Tune 1 is described as follows:

The stressed syllables form a descending scale. Within the last stressed syllable, the pitch of the voice falls to a lower level. (p. 4)

Armstrong and Ward illustrate this Tune 1 with examples like the following:

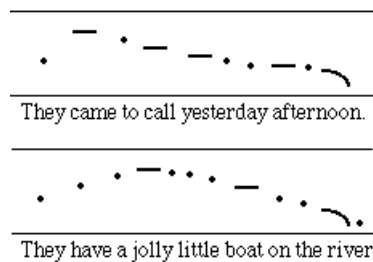


Figure 2. Illustration of the intonation pattern of “Tune 1” from Armstrong and Ward 1922.

In order to transcribe sentences such as these using the INTSINT transcription system (Hirst and Di Cristo this volume) we need to decide whether stressed

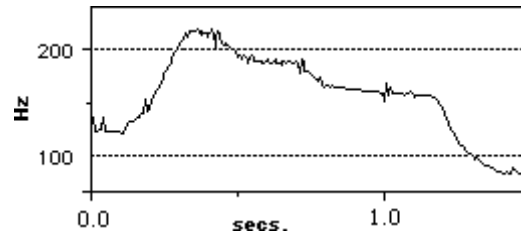


Figure 3. The F_0 curve for the sequence *ma 'ma 'ma 'ma 'ma ma* imitating the intonation of the sentence *But who stole Jane's bracelet?*

extreme case it may become practically imperceptible giving rise to the “hat” or “bridge” type of pattern of sentences (3a) and (3b) that has been described as typical of unemphatic utterances in a number of different languages (see 't Hart this volume and references there).

Whether the downstepped pitch accent is actually suppressed in these cases (as suggested by Knowles 1984 p. 232 and by Hirst 1984 p. 53) or whether the amplitude of the downstep is simply reduced, is, as mentioned above, not always easy to decide from simple observation of F_0 curves. Knowles (1987) makes the same point and claims that:

accent suppression is not all-or-none, it is a process that can apply to a greater or lesser degree. (p. 126)

Reliable empirical criteria will require closer modelling of F_0 curves as well as more work on the psycho-acoustic perceptibility of this type of accent.

Most descriptive studies of English intonation list a variety of different recurrent tonal patterns which can occur in the **head**, *i.e.* the sequence of accents preceding the nucleus. Palmer (1922) lists four varieties of heads: “inferior”, “superior” “scandent” (=sequence of rising accents) and “heterogeneous”. O'Connor and Arnold (1961) identify “low”, “stepping” and “sliding” (=sequence of falling accents) heads although once again this classification is to some extent abandoned in the 1973 revision of their book. The most detailed classification of different patterns was that of Crystal (1969 pp. 225–233) who, on the basis of an extensive corpus study based on a total of about three hours of recording from thirty different speakers, distinguished four categories of globally falling heads, two categories of globally rising heads as well as two mixed categories. The stepping category described above as “neutral non-emphatic” is the most frequent single category of head, occurring in about 30% of Crystal's data. Crystal also notes that whereas “low heads” and “sliding heads” are practically non-existent in his corpus, the two globally rising categories account for another 30% of his data. I shall return to “sliding” heads below (§3.1) in a discussion of intonation patterns of other varieties of English; “scandent” and “globally rising” heads will be discussed under the heading of emphatic variants.

2.2 Mode and expressivity

Armstrong and Ward's second basic tune is said to be used in four essential cases:

- (a) statements with implications
- (b) Yes-No questions
- (c) requests
- (d) incomplete utterances.

They describe the pitch pattern of Tune 2 as follows:

The outline of the first tune is followed until the last stressed syllable is reached. This is on a low note, and any syllables that follow rise from this point. (p. 20)

To anyone outside the field of intonation studies, the very idea that intonation can contribute to the meaning of an utterance is indissociably linked with the distinction between declarative and interrogative intonation patterns distinguished essentially by a falling as against a rising pitch movement at the end of the utterance. Many linguists however have made use of one single rising pattern to describe both continuative and interrogative patterns. When a distinction is made between low rise and high rise, the low rise is generally held to correspond to a statement which is either unfinished or carries implications of some sort, whereas the high rise is said to correspond to a question. Even when such a distinction is made, however, a number of different positions need to be distinguished.

The strongest claim is that the high rise is exclusively used for questions. This seems to be the position of Kingdon (1958b), since, although he does not say so explicitly, in all his examples the high rise is found only on questions and his low rise never occurs on questions except for sequences of alternative questions where the low rise is marked for every group except the last. A slightly weaker position was held by Palmer (1922) who claimed that the high rise could be used both on questions and statements but that the low rise:

is confined to Statements and Commands, it cannot be used for Questions. (p. 84).

A similar position was taken by Halliday (1967a, 1970):

the difference, though gradual, is best regarded as phonetic overlap (...) the one being merely lower than the other (...) But the meanings are fairly distinct. In most cases the speaker is clearly using one of the other; but sometimes one meets an instance which could be either. (p. 21)

By contrast, O'Connor and Arnold (1961) maintain that a low rise is:

by far the most common way of asking Yes/No questions. It should be regarded as the normal way. (p. 55)

but that to turn a statement into a question, a high rise (tone group 8) is needed “as in so many other European languages.” (p. 57). Similarly, Jones (1918) describes a potential distinction between *yes* said with a low rise, meaning *Yes, I understand that, please continue* from *yes* said with a high rise, meaning *Is it really so?* (p. 277). In all the examples of transcription which follow, however, all other interrogative forms are marked with a low rise (pp. 282–283).

A fundamental problem underlying these descriptions is the fact that an utterance which is perceived as a question in a given context may no longer be perceived as a question when taken out of this context. This has been demonstrated experimentally for Edinburgh English (Brown *et al.* 1980) and is almost certainly true also for RP, indeed probably for all languages. The effect of this context dependence is that when asked to produce an interrogative pattern out of context, subjects are liable to produce patterns which may be far less common in spontaneous speech.

A rather different explanation for the distinction between high and low rises was proposed by Cruttenden (1970) who suggested that:

The meaning of a rise ending high which is required to turn *you’re coming* into a question is probably better described as “surprise”. Questions already signalled by the syntax will not usually have high rise but low rise or fall (...). If such questions do have high rise (...) then the element of surprise is added to the question. (p. 188)

It has, in fact, been suggested by several authors (in particular by Bolinger in a number of publications) that it is a mistake to equate the choice of final pitch movement in an utterance with sentence type since, as any study based on utterances produced in spontaneous conditions quickly discovers, it is perfectly possible to find both rising pitch without questions and questions without rising pitch. As Couper-Kuhlen (1986) notes, we need to make a distinction between syntactic sentence type and pragmatic speech act. She goes on to claim that despite the final rise, a sentence such as:

(5) You’ve FInished?
[↓ ↑]

is not a syntactic question since “if it were it would have subject-operator inversion.” This argument, as it stands, appears somewhat circular – it is not sufficient to simply stipulate that subject-operator inversion is necessary for syntactic questions. There is, however, independent evidence, as I have argued elsewhere (Hirst 1983b) that Couper-Kuhlen is right. It is a well-known fact of English syntax that unlike (6a) a sentence such as (6b) is unacceptable with the indicated stressing:

- (6) a He BOUGHT something
- b *He BOUGHT anything.

but that in syntactic questions, both *something* and *anything* are acceptable:

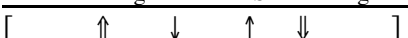
- (7) a Did he BUY something?
 b Did he BUY anything?

The crucial fact is that sentence (6b), unlike (6a) is still unacceptable even when it is provided with rising intonation (or, in a written text, with a question mark):

- (8) a He BOUGHT something?
 b *He BOUGHT anything?

This seems to be conclusive evidence that rising intonation is not, contrary to what has often been claimed, a way of turning a statement into a (syntactic) question, but is rather a way of indicating that a syntactic statement is being used pragmatically as a request for information.

In fact, in spontaneous conversation it is quite possible to produce a syntactic statement with falling intonation which is used as a request for information. The following figure shows the F_0 curve of a sentence (from a published collection of recordings: Dickinson and Mackin 1969 p. 78) taken from a recording of a visit to an optician who asks her patient:

- (9) You MAnage in the DIStance alright?


which is quite clearly intended as a question in the context and is interpreted as such by the patient.

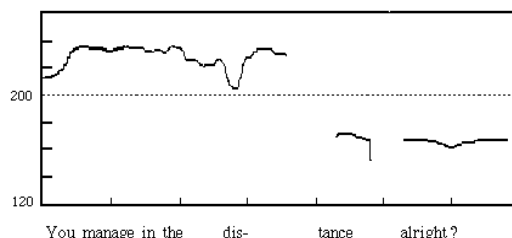


Figure 4. F_0 curve for the "real-life" question *You manage in the distance alright?*

As Lindsey (1985, 1991) points out, questions with declarative form and falling pitch are far more common than is often thought.

We conclude, then, that in English at least, while it is fairly common to use rising intonation for questions, it is by no means compulsory; nor can a rising pattern in itself transform a statement into a syntactic question. We might then ask whether there is some more general pragmatic characteristic which underlies the use of rising pitch both on questions and on statements.

One of the most commonly proposed candidates for such a characteristic has been that of “incompleteness” (cf. Faure 1962 p. 73). Coleman (1914) suggested that Yes/No questions are incomplete alternative questions in which the alternative “or not” has been suppressed and a similar argument has been proposed by a number of linguists although Bolinger (1978a) has convincingly shown that this analysis cannot be correct.

A number of recent proposals, in particular Brazil (1975) and Gussenhoven (1984), have built on an earlier suggestion by Jassem (1952) that:

Falling nuclear tones have proclamatory value. Rising nuclear tones have evocative value. (p. 70)

Under this analysis the basic function of the distinction between falling contours and rising contours is to indicate to the listener how he is intended to process the propositional content of the utterance. A falling contour can be interpreted as an assurance that this propositional content is intended to be added to what **Relevance Theory** (Sperber and Wilson 1986) calls the “mutual cognitive environment” of speaker and listener: the set of facts which at any given moment speaker and listener can share (and can be aware of sharing). This approach seems very promising (see Hirst 1989 for discussion) but much research remains to be done before the numerous insights which have been proposed can be integrated into a formal theory of the pragmatic interpretation of prosody. For a recent attempt cf. Vandepitte (1989). For a similar approach within a somewhat different framework cf. Pierrehumbert and Hirschberg (1990).

2.3 Focalisation and contextual effects

The term **focus**, more specifically **narrow focus**, has in a number of recent works replaced the more traditional term of **emphasis** to refer to the way in which a speaker gives an optional prosodic highlighting to part of an utterance.

Whether we choose to use the term “focus” or “emphasis” may consequently appear to be a trivial question of vocabulary (or fashion). There does, however, appear to be at least one case where the two notions are not entirely identical. Whereas emphasis is basically a **paradigmatic** notion, in that any given item is either emphatic or non-emphatic, focus is basically **syntagmatic** since it applies to one element of a sequence (more strictly to one node of a tree). This implies that while it is quite possible to refer to emphatic and non-emphatic versions of an utterance consisting of a single word, (such as “NO.” vs. “NO!”), it is not possible, in standard treatments of focus, to distinguish broad and narrow focus on a single item. The choice between the two notions would then boil down to answering the empirical question: is there a categorical distinction between emphatic and non-emphatic readings of a single word? I assume that

such a distinction does exist in British English and shall consequently prefer the concept of emphasis to that of focus.

Classical descriptions of English intonation, since Coleman (1914), refer to two types of emphasis: **contrast emphasis** and **intensity emphasis**. Jones (1918) remarks that:

Contrast emphasis may be applied to almost any word, but intensity emphasis can only be applied to certain words expressing qualities which are measurable. (p. 298)

Intensity emphasis is simpler to define in that it is semantically approximately equivalent to adding an intensifying adverb such as “absolutely”. Thus (10a) with intensity emphasis has approximately the same interpretation as (10b):

- (10) a This chocolate is delicious.
b This chocolate is absolutely delicious.

When intensity emphasis is applied to the first auxiliary of an utterance the result is the equivalent of an exclamative formed with *How* or *What* so that (11a) is equivalent to (11b):

- (11) a This chocolate is delicious!
b How delicious this chocolate is!

The two types of emphasis described above obviously have much in common. In particular, the final (nuclear) pitch accent typically rises to a higher level than that of the preceding syllable, unlike the pattern described in §2.1 for unemphatic utterances. A minimal emphatic reading of (10a) (whether contrastive or intensive) would consequently be:

- (12) a This CHOcolate is deLlicious.
[↑↑ > ↑ ↓]

Very often, the effect of the high-falling final pitch accent is reinforced by a low-pitched initial accent, itself often preceded by high-pitched unstressed syllables giving a rising head:

- (12) b This CHOcolate is deLlicious.
[↑↑ ↓ ↑ ↓]

The difference between the low initial accent in (12b) and the high initial accent as in (12a) is a qualitative difference, unlike the difference between a high-falling and a low-falling nucleus which is simply a question of degree. It is perhaps for this reason that this is a very common way of signalling emphasis. Interestingly, the same pattern can occur on a single word containing more than one stress as in:

(13) $\frac{\text{perCEPtIBility}}{[\uparrow \downarrow \uparrow \downarrow]}$

The fact that a single word can be pronounced with a pattern of this type is further evidence that, as mentioned above, it is a paradigmatic opposition which is involved here rather than a syntagmatic contrast.

The pattern consisting of a rising head followed by a high-falling nucleus can be used to express other types of meanings which cannot all be put down to intensive or contrastive emphasis. A similar contour is discussed by Liberman and Sag (1974) and Liberman (1975) for American English with an explicit comparison with British English patterns of the same form. These authors assume that the pattern expresses a global meaning which they call **surprise/redundancy**. In many cases, if not all, the meaning could just as appropriately be labelled **exclamative**, a label which, as we saw above, can also be applied to most of the cases described as “intensity emphasis”.

On a longer sentence, any stressed syllables between the initial low accent and the final high-falling accent will be on an intermediate pitch. Just as with the downstepping head described in §2.1, any accented syllables in a rising head of this type are liable to be signalled by a flattening out of the F_0 curve on the accented syllable, and will be consequently transcribed as in (14):

(14) $\frac{\text{It's ONE of the MOST auTHENTIC PUBS in SUsex.}}{[\uparrow \downarrow < < < \uparrow \downarrow]}$

Upstepping accents of this sort have not always been considered genuine pitch accents. Thus Cruttenden (1986) gives an example which he transcribes:

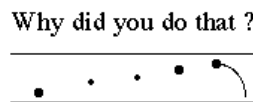


Figure 5. Schematic illustration of a rising head (from Cruttenden 1986).

Cruttenden states that whereas **Why** and **that** carry pitch accents, **do** has only tertiary stress (dependent on length and/or loudness alone) (p. 56). It is of course an empirical question whether in any given utterance an upstepping pitch accent is actually present (once more a subject for further research!) but in the affirmative I should transcribe Cruttenden’s example as:

(15) $\frac{\text{WHY did you DO THAT?}}{[\downarrow < \uparrow \downarrow]}$

A further variant of the rising head (O’Connor and Arnold (1969 2nd edition 1973) call this the “climbing head”) consists of a recurrent sequence of local rises from each accented syllable giving the following pattern:

- (16) It's ONE of the MOST auTHENtic PUBS in SUsssex.
 [↑ ↓ ↑ ↓ ↑ ↓ ↑ ↓ ↑ ↑ ↓]

The way in which emphasis interacts with the basic opposition between rising and falling tones is, once again, a subject of some controversy. Armstrong and Ward (1926) considered that the falling-rising nucleus was the emphatic form of the rising nucleus. This analysis has been implicitly accepted by many authors although others have explicitly rejected it. Halliday (1967a), for example, claimed that there are five basic nuclear tones: falling, high rising, low rising, falling-rising and rising-falling, none of which can be considered modifications of other tones and most of which have secondary variants. Brazil (1975), claimed that it is the fall-rise which is basic and that the simple rising pattern is an emphatic modification of the fall-rise. Gussenhoven (1984) has claimed that there are three primary tones: falling, rising and falling-rising, each of which can exhibit secondary modifications.

It is not at all obvious how one might hope to settle a controversy of this sort other than by appealing to theory-internal criteria. One possible direction is that taken by Lindsay and Ainsworth (1985) who attempt to show that a five-tone analysis makes better predictions than a two-tone analysis concerning discrimination functions derived from listeners' judgements on synthetic stimuli. Another type of experiment was conducted by Gussenhoven (op. cit. chapter 7) who asked listeners to judge the similarity/difference of synthetic stimuli. Gussenhoven interprets his results as evidence against Brazil's analysis. Both of these experiments, however, suffer from weaknesses in the choice of stimuli as well as in their reliance on synthetic speech. Neither can be considered sufficiently convincing to constitute reliable proof of one analysis or the other. Both, however, constitute extremely interesting attempts to provide much needed empirical evidence for phonological analyses.

2.4 Phrasing and textual organisation

Crystal (1969) noted that the average length of Intonation Units in his corpus was of five words and that 80% of the Units were less than eight words long. When utterances are longer than this, they are usually broken up into two or more Intonation Units. WH-questions appear to impose greater restrictions on the possible intonation breaks during the utterance so that a long question will still tend to be produced as a single Intonation Unit even in a sentence containing as many as eight accents:

- (17) WHAT MADE JOHN TELL ANNE NOT to GO HOME?
 [↑ > > > > > > > ↓]

Apart from this type of question, it is fairly rare to find utterances in spontaneous speech which contain more than three or four pitch accents in a single Intonation Unit.

There has been considerable disagreement as to what criteria, syntactic, semantic or pragmatic, are relevant for this phrasing. For a summary of arguments for and against syntactic constraints on phrasing cf. Couper-Kuhlen (1986 especially chapter VIII). Many of the arguments which have been presented against such constraints, however, no longer hold if we assume a less trivial correspondence between syntax and phonology than has generally been proposed. Thus it has generally been supposed that a grammatical account of phrasing must show a one-one correspondence between syntactic units and prosodic units. This is obviously not the case in utterances like the following (from Couper-Kuhlen 1986; the symbol / indicates the observed boundaries):

- (18)a /They feel like they're a forgotten bit / of a war / that nobody wants to solve /
b /They'll leave it alone / till it splatters out / to a deadly end/
c /So here I am / in the middle of the most enormous / movement /
d /as if the whole world / is hanging waiting on our decision /
e /which I found one of the most fascinating and most interesting / times of my life/

Couper-Kuhlen consequently draws the conclusion:

it is virtually impossible to predict where boundaries will come. (p. 153)

I have suggested (Hirst 1987, 1993) an alternative explanation for this apparent lack of correspondence between syntactic and phonological constituents. While pragmatic and phonological constraints are obviously the ultimate criteria by which a speaker decides where he **will** place a boundary, syntactic criteria define where these boundaries **may** occur. In all the examples in (18), as well as in others given by the same author, it is striking that each boundary occurs **before** a complete syntactic constituent extending to the end of the sentence. The reason why the correspondence between syntactic and prosodic constituents breaks down is that syntactic constituents may be interrupted by a prosodic boundary at the beginning of an internal syntactic constituent provided that a prosodic boundary is also placed at the end of that constituent. Thus in (18) for example, the syntactic structure relevant to the phrasings noted is:

- (19)a [They feel like they're a forgotten bit [of a war [that nobody wants to solve]]]
b [They'll leave it alone [till it splatters out [to a deadly end]]]
c [So here I am [in the middle of the most enormous [movement]]]
d [as if the whole world [is hanging waiting on our decision]]
e [which I found one of the most fascinating and most interesting [times of my life]]

This interpretation thus predicts that while several different phrasings may be theoretically possible, many others will be ruled out; in particular internal boundaries are predicted not to occur before a constituent the end of which is not also marked by a boundary.

While there is, as we saw above, quite a remarkable consensus concerning the existence of prosodic constituents equivalent to what I have called Intonation Units, there is considerably less agreement as to whether larger prosodic units need to be identified.

It has been suggested that Intonation Units are organised into higher-order “paratone-groups” (Fox 1973, 1984) or “major paratones” (Yule 1980), which are signalled essentially by a change of overall width of pitch range or “key” (Brazil 1975, Brown *et al.* 1980).

The beginning of a paratone is said to be usually marked by extra high pitch on the first accent while the end is usually marked with extra-low pitch. When the end of a paratone is marked in this way but not the beginning, the result is what Yule has called a “minor paratone”. It seems however equally possible to mark the beginning of a paratone but not the end. This suggests that rather than distinguish major and minor paratones we might make a four-way distinction between Intonation Units which are marked as paratone-initial, paratone-final, or both or neither. Such a distinction could be marked in an INTSINT transcription by simply doubling the initial or final square bracket of an intonation unit so that it would be possible to have a sequence such as:

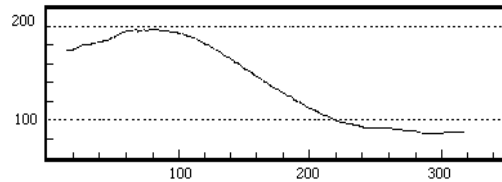
(20) [[A] [B] [[C] [D]] [E]]

where A and C are marked as paratone-initial and D and E as paratone-final, even though the sequence as a whole is not properly bracketed and cannot be divided into a sequence of independent paratones.

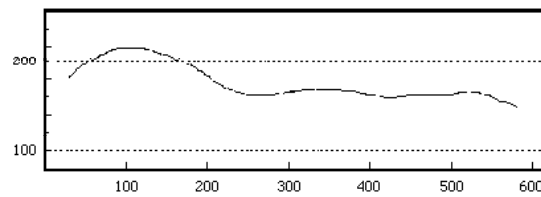
Another line of research concerns the prosodic structure of conversation and the role of pitch in “turn-taking” (Cutler and Pearson 1986) and “interruption-management” (French and Local 1986). Both of these fields obviously need to be developed (for an overview see Couper-Kuhlen 1986 chapter XI) although on the basis of preliminary results it does not seem very likely that the various strategies described will show many language specific characteristics which need to be accounted for in the phonological description of a given language.

2.5 Stylised patterns

Like many languages, English makes use of a certain number of patterns which strike the listener as being somehow intermediate between speech and song. The common prosodic characteristic of these patterns is that instead of consisting of a continuous sequence of movements from one target-point to the next as in normal speech, the contour is produced as a sequence of static level tones.



a. John!



b. Jo - ohn!

Figure 6. F_0 curve for the non-stylised contour (a) and the stylised contour (b) used for calling. Horizontal lines correspond to 100 and 200 Hz.

The semantic effect of these contours has been discussed in detail by Ladd (1978) who has aptly called them **stylised** patterns, the common feature being one of stereotyped, conventional almost ritual behaviour. It has often been noted that these patterns are particularly frequent in children's speech, particularly in jeering chants like

(21) $\frac{\text{MOLly is a BA- by}}{[\rightarrow \downarrow \uparrow > >]}$

Stylised patterns which are common in adult speech are vocatives ("Jo-ohn!"), and greetings (Good Morning!) particularly in situations such as answering the phone, for example, where the speaker repeats the same message many times throughout the day.

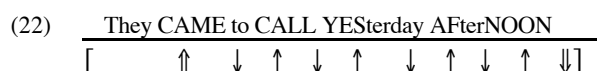
3. Comparisons with other systems

3.1 Comparisons within the same language

British English, as mentioned in §1, is spoken with a number of different accents, some of which exhibit quite strikingly specific intonation variants, concerning both the recurrent patterns found on the head as well as the pattern occurring at the nucleus. Unfortunately, studies of these regional characteristics are few and far between.

a. Recurrent patterns

The downstepping pattern described in §2.1(c) as typical of unemphatic statements is replaced in a number of dialects by a sequence of falling pitch accents (the “sliding head” mentioned above) so that instead of (4a) we find:



This has been described as typical of Scottish accents, both Western (McClure 1980) and Edinburgh (Brown *et al.* 1980). Neither Palmer (1922) nor O’Connor and Arnold (1961) nor Crystal (1969) describe this as a variant of the unemphatic pattern for Standard British. According to O’Connor and Arnold the sliding head is only found before a falling-rising nucleus. The pattern is probably gaining ground throughout England possibly due to the influence of American speech where the pattern is very common (cf. Pike 1945, Bolinger this volume).

A sequence of rising pitch accents has been described by a number of authors as typical of Welsh accents although I am unaware of any detailed study of the intonation of Welsh English.

b. Nuclear patterns

English accents from Northern Britain, particularly Belfast (Cruttenden and Jarman 1976), Liverpool (Knowles 1974, 1978), Birmingham, Glasgow (Brown, Currie and Kenworthy 1980) and Tyneside (Pellowe and Jones 1978, Local 1986)) are notorious for the fact that, unlike what has been described for most languages, they commonly make use of an intonation pattern with high or rising final pitch in what native speakers perceive as perfectly ordinary statements. For recent discussion cf. Knowles 1984, Cruttenden 1986, Bolinger 1989). Knowles suggested that these rising pitch movements should in fact be interpreted as falls (he calls them “Irish” falls which, perversely, go up) and concludes that the existence of such pitch contours shows up a major weakness in most current systems of analysis which classify patterns according to predetermined phonetic characteristics. An comprehensive account of these rising patterns is to be found in Cruttenden (1994) together with a comparison with superficially similar but functionally different patterns found in an area which Cruttenden refers to as the “pacific rim” (*i.e.* West Coast USA, Australia *etc.*). (see also Bolinger this volume). It has been suggested that the British pattern is of Celtic origin which would explain some of the distribution in the West of the British Isles. This would not however explain why the pattern is far less common in Eire than in Northern Ireland, nor why it is to be found in the Newcastle area. An intriguing possibility would be that this pattern is in fact a trace of the Viking occupation of Britain – similar patterns have been described

for East Norwegian (Fretheim and Nilsen 1989) and West Swedish (Gårding this volume).

**4. Theoretical implications and conclusions:
the phonology of English intonation**

One of the basic aims of phonological theory is to attempt to explain how and why languages differ from one another phonetically, by setting up a limited number of phonological parameters which can combine in various ways to generate the appropriate range of phonetic variability.

This section sketches a theory of phonological representations of intonation which attempts to account for some of these parameters as well as some of the dialectal variations mentioned above.

I assume here, as proposed in the framework of autosegmental phonology (Goldsmith 1976, 1990) that phonological representations in all languages consist of two distinct lines of phonetically interpreted segments: tonal segments (or tones) and phonematic segments (or phones). I also assume, contrary to standard autosegmental theory, but following Hirst 1983a, 1984 (see also Pierrehumbert and Beckman 1988) that these two lines are related to each other **indirectly** via a hierarchical structure containing at least two levels of constituents: Tonal Units and Intonation Units.

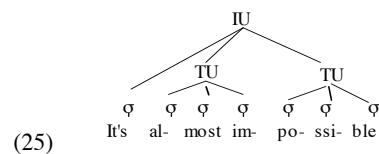
Since English is not a tone language, tonal segments are not specified in the lexical entry for each word but are rather added to the phonological representation respecting a phonological template which for English could be of the form:



If we assume that the appropriate prosodic structure for a sentence such as:

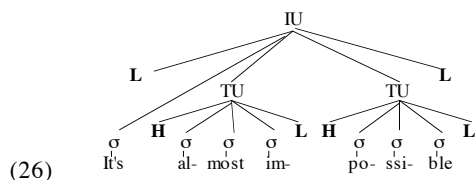
(24) It's almost impossible.

is:

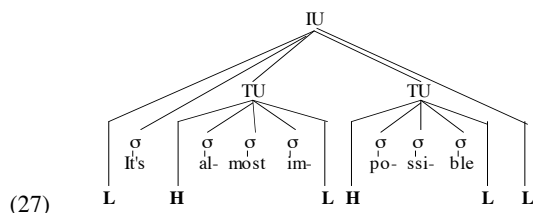


(where σ represents the syllable constituent) then, in order to respect the tonal templates (23a,b), tonal segments will need to be added. Assuming for example that terminal intonation is chosen, this will give:

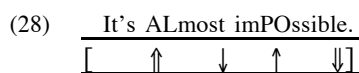
British English



Such a representation, however, is not pronounceable, since the tonal segments are only partially ordered. Total ordering could be achieved either by reassociating the tones assigned to the Intonation Unit to Tonal Units (as suggested in Hirst 1986) or, perhaps more appropriately, following Pierrehumbert and Beckman (1988) by assuming that tonal segments are all projected onto the same tier but remain linked to different hierarchical constituents.¹ This would result in:

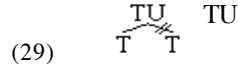


The advantage of such a representation is that information concerning the hierarchical level of the constituent to which a tonal segment is attached is available to the rules of phonetic interpretation which convert (27) into something like:



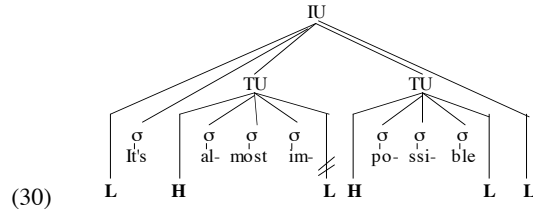
I shall not go into any more detail here concerning the phonetic rules which are assumed to derive a representation such as (28) from (27). Note simply that two consecutive L tones may be interpreted as a single low phonetic target pitch.

The intonation contour generated in (27) is not that described above as the basic unemphatic pattern for British English, but rather that which is described as basic for American and Scottish dialects (§3.1). In order to derive the British English pattern we need to assume a further parameter converting a sequence of falls into a downstepping pattern. In lexical tone languages, a downstepped tone is standardly analysed (following Clements and Ford 1979) as a high tone preceded by a “floating” low tone – *i.e.* a low tone which is not phonetically realised as a pitch target but which has the effect of lowering the following high tone. This can be achieved very simply by a further rule specific to British English which “delinks” the second of two linked tones in all but the last Tonal Unit:

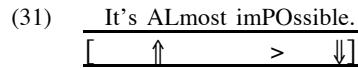


where **T** is either **H** or **L**

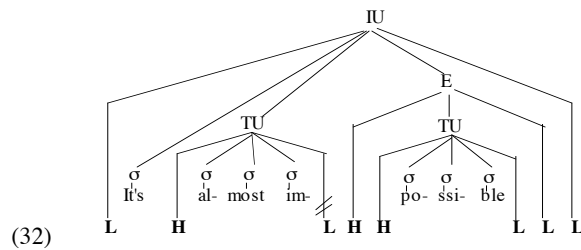
Applied to (27) this will result in:



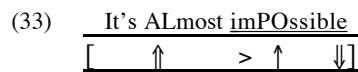
which is then interpreted as:



The emphatic patterns discussed in (§2.4) could be derived in a number of ways. I have suggested (Hirst 1983b) that these contours contain an emphatic morpheme consisting of a single floating High tone. Another possibility would be to assume that a new prosodic constituent **E** is introduced into the prosodic structure between the Intonation Unit and the Tonal Units and that this constituent is assigned the same tonal segments as the Tonal Unit. This would result, after delinking of the low tone in the first Tonal Unit, in the following structure:

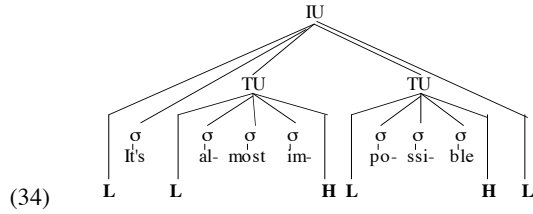


which can be interpreted as:

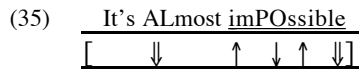


The emphatic patterns with “rising head” (= sequence of upstepped accents) and “climbing head” (= sequence of rising pitch accents) suggest the possibility that an alternative template for the Tonal Unit is available with the sequence [L H] instead of [H L]. Applied to (25) this would result in:

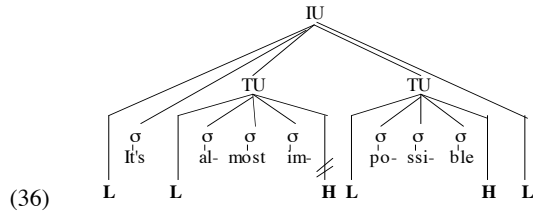
British English



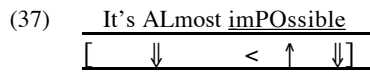
interpreted as:



or alternatively, if (29) applied, in:



interpreted as:



A number of questions remain unanswered concerning both the way in which phonological representations of intonation structures are derived and the way in which these representations are interpreted phonetically. The brief outline given in this section, however, does seem to possess at least some of the characteristics of what a phonological theory of intonation might look like, as well as how such a theory might apply to generate the observed variety of patterns in British English.

Note

1 Note that Pierrehumbert and Beckman (1988) assume in addition that tonal segments can be multiply linked to constituents of the hierarchical structure, a suggestion which I do not follow.