7

Tonal Structures

7.1 Introduction

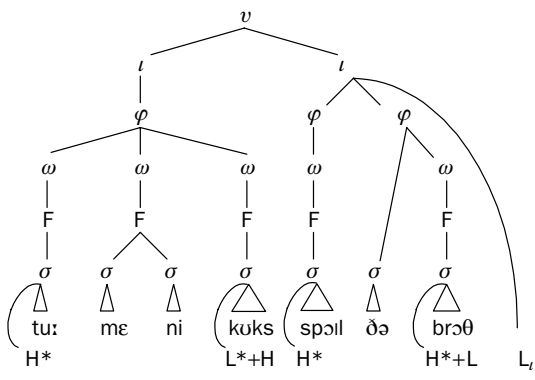
This chapter offers an account of Pierrehumbert's model for the description of tone and intonation. Her seminal thesis (Pierrehumbert 1980) presented a descriptive framework for intonation which separated the phonological representation from its phonetic implementation. This made it possible to characterize the notion 'possible prosodic structure' independently of the phonetic details of intonation contours. Section 7.2 places the model in its historical setting by identifying the positions in the pre-1980 literature that are intellectually closest to its various elements. The section ends with a nutshell description of the revised grammar of American English of Beckman and Pierrehumbert (1986). Section 7.3 discusses the developments that have taken place since the revised 1986 model was presented in Beckman and Pierrehumbert (1986). Tonal structure is sensitive to prosodic phrasing because prosodic phrases may begin and end with boundary tones. Section 7.4 looks at a second effect of phrasing on tone structure, that of the rhythmic distribution of pitch accents, or the resolution of *stress clash.*

Earlier introductions to intonational phonology are Ladd (1996), who coined the term 'Autosegmental-Metrical (AM) model' for it, Shattuck-Hufnagel and Turk (1996), and Beckman (1996). The model is autosegmental because it has separate tiers for segments (vowels and consonants) and tones (H,L). It is metrical because it assumes that the elements in these tiers are contained in a hierarchically organized set of phonological constituents, as depicted in (1), to which the tones make reference in several ways. The tones, which are organized into pitch accents and boundary tones, may or may not be associated with TBUs. Representation (1) gives a possible pronunciation of the English proverb *Too many cooks spoil the broth.* The example illustrates a commonly adopted set of prosodic constituents for English. Among these is the intonational phrase (ι), which has initial and final boundary tones (T_ι). A non-crucial assumption is that in the analysis of English intonation to be presented in chapter 15, final boundary tones are optional, the

first ι in (1) not having any. As is clear from (1) English has both monotonal and bitonal pitch accents, the starred tone of which associates with the syllable. Instead of these graphically elaborate associations, pitch accents are shown as associating with the vowels in the accented syllable, and boundary tones with the constituent brackets, a notation introduced by Hayes and Lahiri (1991b) (but see section 8.3.4).

As explained in chapter 4, the surface representation (1) is a theory of a mental construct for a speaker of English. This speaker also possesses a phonetic implementation module which, among other things, translates every tone into an F_0 target. The phonetic 'alignment' with the segmental tier, its timing, will to some extent be language-specific, as will its 'scaling', its F_0. As explained in chapter 5, over and above the effects of these implementation rules, the speaker's psychological condition and communicative purpose will influence the overall pitch range.

(1)



This chapter is concerned with the phonological structure. However, a complete analysis of an intonational system will comprise a phonology as well as a morphology. Unlike the tones in the phonology, the elements in a morphological analysis are not given *a priori*. In principle, the ι-wide contour could be a single morpheme, or any single tone or sequence of tones contained within ι could be a morpheme. A whole-contour proposal occurs in Liberman and Sag (1974), where the 'contradiction contour', for example, consists of a two-accent sequence, a position which has not been followed (Bolinger 1986: 245). Morphemes consisting of pitch accents plus boundary tones occur in proposals' intonational meaning in English (Brazil, Coulthard, and Johns 1980; Gussenhoven 1983b; Cruttenden 1997) and Bengali (Hayes and Lahiri 1991a). Pierrehumbert and Hirschberg (1990) propose an analysis in which every tone is a morpheme, a position defended in Bartels (1997). Dainora (2001, 2002) advanced the argument that the relatively high predictability of transitions between pitch accents and following tones in American English suggests that these elements are not morphemically independent. Of course, the same phonological analysis may be

compatible with a large number of morphological analyses, just as different views of the morphological structure of a word like *replicate* ([[*re* [*plic*]] *ate*], [[*replic*] *ate*], [*replicate*]?) do not compromise the assumption that its surface phonology is ['rɛplɪˌkeɪt].

7.2 Historical background

The description of American English given in Pierrehumbert (1980) was somewhat revised in Beckman and Pierrehumbert (1986), while the model was consolidated as a general theory of intonational representation in Pierrehumbert and Beckman (1988). In this section, a number of aspects are isolated and related to earlier positions.

7.2.1 Autosegmental-Metrical representation

The idea that there are separate tiers for tones and segmental phonemes was implicit in many descriptions of intonation, inasmuch as pitch features were not considered to be part of the featural composition of segments in the British tradition of intonation description (e.g. O'Connor and Arnold (1973) or the description of Dutch by 't Hart, Collier, and Cohen (1990)). Also, Goldsmith's (1976) autosegmental model had been applied to English intonation before by Goldsmith (1980) (which began as an unpublished MIT paper in 1974), Liberman (1975), and Leben (1975).

The idea that at least one prosodic constituent, the ι, was intonationally defined, had been current in prosodic research at least since Selkirk (1978), and figured in earlier descriptions in the form of the 'tone group' (e.g. O'Connor and Arnold 1973). Prosodic phonology holds that speech is produced in batches of segments that are hierarchically ordered: within any such batch except the lowest a smaller batch can be identified. The gestural integration correlates inversely with rank: syllables are highly integrated articulations, while Utterances may contain noticeable pauses. A widely adopted view of this hierarchy for English includes the syllable (σ), the foot (F), the phonological word (ω, also 'prosodic word'), the phonological phrase (ϕ), the intonational phrase (ι), and the Utterance (υ) (Selkirk 1978; Nespor and Vogel 1986; Hayes 1989). This hierarchy is illustrated in (1).

In the revised theory of Beckman and Pierrehumbert (1986), an additional intonationally defined constituent was introduced for English, the Intermediate Phrase (ip), ranked immediately below ι. This made two degrees of depth available for an intonational boundary. Examples of 'lower' boundaries given by Beckman and Pierrehumbert (1986) are illustrated in (2) and (3), where the square brackets enclose ips and the curly brackets the ι. In (2), the two adjectives are considered to be followed by just an ip-boundary, because the disjuncture with what follows is less complete than that observed for a fully-fledged ι-boundary. In (3), the

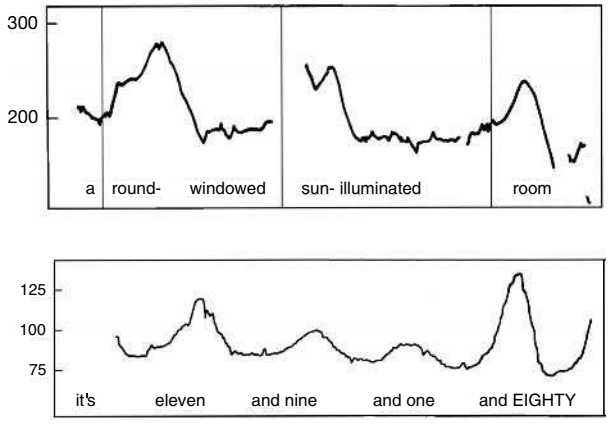


Fig. 7.1 F_0 contours of (2) and (3). From Beckman and Pierrehumbert (1986)

ip-boundary after *nine* is motivated by the high F_0 peak on *eighty*, which is due to an interruption of the downstepping pattern shown by the preceding F_0 peaks on *one* and *nine*. The F_0 contours are given in panels (a) and (b) of figure 7.1, respectively.

(2) { [A round-windowed] [sun-illuminated] [room] }

(3) { [It's eleven and nine and one] [and eighty] }

However, there are many analyses of West Germanic languages, including the description of English in chapters 14 and 15, that do without ip and the 'phrase tone' that has been related to it (van den Berg, Gussenhoven, and Rietveld 1992; Féry 1993; Grabe 1998a; Gussenhoven, Terken, and Rietveld 1999). In those analyses, the internal ip-boundary of (2) would be an ι-boundary, while that in (3) would not be an intonational boundary at all; the exemption from downstep is attributed to fact that it is the last ('nuclear') pitch accent of the ι (Ladd 1983c; Gussenhoven 1983b; Truckenbrodt 2002a).

Work on Japanese, Korean, and Basque led to the inclusion of the Accentual Phrase (α) in those languages, ranked immediately above the ω (Pierrehumbert and Beckman 1988; Jun 1998; Elordieta 1997) and comparable to ϕ (see also chapters 9 and 10).

7.2.2 Pitch accents and boundary tones

The separation into pitch accents and boundary tones harks back to Trager and Smith Jr's (1951) juncture phonemes (# 'falling', – 'sustained', || 'rising'), which existed by the side of the pitch phonemes. A early division between pitch accents and boundary tones can also be recognized in the work in the 1970s and 1980s by Hans't Hart, René Collier, and Antonie Cohen on Dutch ('t Hart, Collier, and

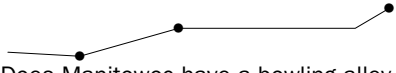
Cohen 1990), as argued by Ladd (1996). They divided up the pitch movements of Dutch into 'accent-lending' and 'non-accent-lending', but stopped short of equating the latter with boundary features. An example is given in (4), a non-final contour, where '1' is an accent-lending rise, 'B' a non-accent-lending fall, 'A' a steep accent-lending fall, and '2' a non-accent-lending rise. (When two accent-lending movements appear on the same syllable, only one accent is produced.)

(4)

We gaan vanavond naar de schouwburg
'We are going to the theatre this evening,...'

The chief motivation in Pierrehumbert (1980) for ending ι with two tones, a phrase accent (T-) and a boundary tone (T%), was that in many contours two targets can be identified after the last ('nuclear') pitch accent. For instance, in (5), the L* pitch accent is followed by a high target at the end of the accented *'Manitowoc*, as well as a final high target, for which H- and H% were postulated, respectively. I reproduce her contour 2.29 as (5). This example would be natural in a conversation where someone had just asked the speaker if he knew of any towns with bowling alleys. In Beckman and Pierrehumbert (1986), T- was reanalysed as a boundary tone of the ip. As a result, ips ended in T-, and ιs in T-T% in the new analysis, since the right edge of every ι coincides with that of an ip.

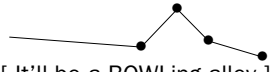
(5)



{ [Does Manitowoc have a bowling alley] }
L* H- H%

The decision to postulate two tones after the pitch accent of course depends on the analysis of the tone string into pitch accents and boundary tones. One of the most striking differences between the Pierrehumbert model and the earlier descriptions of British English (Crystal 1969; Halliday 1970; O'Connor and Arnold 1973) is that Pierrehumbert analysed the contour leading *towards* the accentual target as the pitch accent, while the British tradition isolated the part leading *off* it. Thus, in (5), the contour described by L*HH% is a unit, the nuclear tone '(low–to–)high rise'. This difference in approach is even clearer in the case of contours like (6): for O'Connor and Arnold, for instance, this contour consists of a (high) fall, preceded by a low 'prehead' (see (6a)), between which the pitch will necessarily rise; Pierrehumbert describes it as a rising pitch accent followed by the boundary tones L- and L%, as in (6b).

(6)



{ [It'll be a BOWLing alley] }
a. low prehead | high fall
b. L+H* L- L%

The British 'off-ramp' analysis was the basis of my own autosegmental description of English and Dutch (Gussenhoven 1983b; Gussenhoven 1988; van den Berg, Gussenhoven, and Rietveld 1992), which analysed (5) as L*H H% and (6) as H*L L%. Pierrehumbert's 'on-ramp' analysis was transferred to the practically oriented transcription system for American English that was derived from the Pierrehumbert model, *Tones and Break Indices*, ToBI (cf. Beckman and Ayers 1994, see chapter 15). This issue has not attracted any empirical research, as far as I am aware (see further section 7.3.1).

7.2.3 Associated and unassociated tones

The Africanists' distinction between associated and unassociated lexical tones was applied by Pierrehumbert (1980) to intonational tones in two ways. The first use of 'floating' tones is based on the role that such tones have as triggers for downstep while remaining without a target themselves (see chapter 6, section 6.3.1). Thus, Pierrehumbert (1980) describes one of the downstepped patterns of English as H*+L H*, whereby the trailing L-tone causes the second H* to be downstepped, without acquiring a low target itself. The non-realization of L in H*+L needs to be specified in the implementation rules, and is thus specific to Pierrehumbert's description of American English.

The second, more generally accepted, use is for tones whose targets are timed with reference to the targets of other tones, rather than to a specific point in the segmental tier. Typically, one tone in a pitch accent, T*, associates with the accented syllable, while the leading or trailing T receives a target which occurs some fixed distance before or after that of T*. As a result, the H of L*+H as used on ˈ*Rigamarole* will occur on *-ma-*, the second syllable after the accented *Rig-*, while in *Stein*, it will be realized on the same syllable as L* (Beckman and Pierrehumbert 1986). In chapter 11, section 11.3, we will see that the timing of the Swedish focus-marking tone depends on the association of preceding lexical tones in the word.

7.2.4 Targets and interpolations

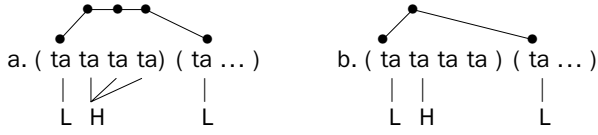
The notion that a contour is an interpolation between points, such that only the beginning and end-points of a movement result from a tonal specification, was inherent in the analyses by Pike (1945) and Trager and Smith Jr (1951), who used numbers to indicate four pitch 'levels'. Pike's (7), for instance, indicates interpolations by means of dashes. In his system, '1' was the highest pitch phoneme, '4' the lowest. The notion of a starred tone was present in that accented levels, which begin a 'primary contour', are marked with the degree symbol. In this case, the representation translates quite readily into Pierrehumbert's (8).

(7) I wanted to do it, but I couldn't
4- °2- -4-3/4- °2- -4 //

(8) I wanted to do it, but I couldn't
H* L-H% H* L-L%

From Pierrehumbert (1980) onwards, a distinction has generally been made between interpolation, the creation of phonetic values between the phonetic targets of phonological tones, and specification through spreading. Spreading is illustrated in example (9a), which shows an H-tone spreading to the end of its phrase, from where the contour continues to the initial L of the next phrase. By contrast, in (9b), the H-tone does not spread, and the interpolation thus includes the unspecified syllables before the phrase-initial L. The *locus classicus* is Beckman and Pierrehumbert (1986: 263) (also Pierrehumbert and Beckman 1988: 37ff.), who showed that in a case exactly like this, (9b) is the superior theory for Japanese. This representation predicts that the longer the first phrase is, the less steep is its sloping F_0. Representation (9a) would incorrectly predict that the slope remains high up till the last syllable of the first phrase, regardless of its length. In chapter 10, section 10.7, this case is dealt more fully. The notion of phonetic underspecification was subsequently extended to other features, like nasality (e.g. Cohn 1990).

(9)



Other theories identified pitch *movements* as the basic elements instead of the interpolations between level pitches, although the idea of non-specification was present inasmuch as they contained stretchable movements whose duration was determined by the length of the segments over which they were pronounced, like the 'rising head' of O'Connor and Arnold (1973) or the gradually rising pitch movement '4' of 't Hart, Collier, and Cohen (1990).

7.2.5 Lexical and intonational tones in a single tier

The idea that the string of tones contained lexical and intonational tones forms the hallmark of Bruce (1977), who isolated the contribution of the lexical tones of Stockholm Swedish from that of the intonational tones, representing them as a string of pitch levels that were timed with the stresses and phrase ends much as in an autosegmental description (Pierrehumbert 2000; Ladd 2000). Ladd (1983b) characterized this type of description as the Tone Sequence model, to distinguish it from descriptions that superimpose accentual contours on phrasal intonation contours, termed Contour Interaction models by Ladd, as represented by Gårding (1983), Thorsen (1978, 1983), and Vaissièrre (1983), as well as by Fujisaki's model (e.g. Fujisaki 1983).

Interactions between lexical and intonational tones are common in tone languages, where initial or final boundary tones may affect the values of adjacent lexical tones. Processes that refer indiscriminately to lexical and intonational tones strongly support the position that these tones form a single phonological string. For instance, in Mankon, H is upstepped after a floating L preceded by LH. The first L in the LHⓁH sequence can be a lexical tone, as in (10a), or an intonational boundary L_ι, as in (10b) (Hyman 1993). The presence of the first L is crucial, since without it, the H *before* Ⓛ is upstepped. Examples of such interactions in Dutch tonal dialects are given in chapter 12.

(10) a. { bɨ? ɨ sinɔ
| | |
L HⓁH
'termite of bird'

b. { bɨ yiŋɔ }
| | | |
L_ι HⓁH L
'they have come'

7.2.6 Only two tones

The idea that only two tones, H and L, suffice to describe languages with intonation and some lexical tone was part of Bruce's thesis. Earlier, Liberman (1975) had described the intonation of American English with the help of four tones, raised H, H, raised L, and L, thus staying closer to the older descriptions. At that point, descriptions were still vulnerable to Bolinger's (1951) criticism that four-level transcriptions of English intonation, like (7), were arbitrary, because 2–4 would not be discretely different from, say, 3–4 or 1–4. But perhaps more so than Bruce's (1977) 'pitch rules', which filled in details like the copying of F_0-values through level parts of the contour, Pierrehumbert's implementation rules made it clear that the generation of an infinite number of F_0 values between the highest and lowest pitches allowed the distinction between phonological representations and phonetic contours to become particularly clear. To characterize the contrasts of the language and thus the notion of a possible intonational structure, two tones suffice, for English and for many other languages that have been described since.

7.2.7 The 1986 model

Beckman and Pierrehumbert's (1986) revised model includes six pitch accents, H*, L*, H*+L, L*+H, L+H*, and H+L*. An optional initial boundary %H precedes the ι for high initial pitch, mid and low beginnings being tonally unspecified. The ip and ι are obligatorily closed by boundary tones, as explained above. The grammar can be given as in (11), where parentheses include optional elements, accolades alternative options, and subscripts stand for '*n* or more occurrences', as usual. The part enclosed between the outermost $(. \ . \ .)_1$ indicates the ip, of which there must be one or more, containing one or more pitch accents.

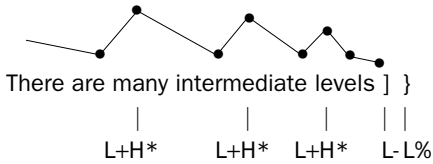
(11) The tonal grammar of Beckman and Pierrehumbert (1986)

$$(\%H)\ \left(\left(\left\{\begin{array}{l}H^*\\ L^*\\ H^*+L^*\\ L^*+H^*\\ L+H^*\\ H+L^*\end{array}\right\}\right)_1\ \left\{\begin{array}{l}H^-\\ L^-\end{array}\right\}\right)_1\ \left\{\begin{array}{l}H\%\\ L\%\end{array}\right\}$$

The phonetic implementation rules include downstep, which applies to H* and H- after a bitonal pitch accent, expressed in (13). A mid tone at the end of an ι is obtained by the combined working of DOWNSTEP (lowering H- after L+H*) and UPSTEP (14), which raises L% to the level of a preceding H-, and raises H% above the level of a preceding H-. Thus, in (15a), H- is downstepped and L% is upstepped. UPSTEP is also responsible for the extra-high H% in (15b).

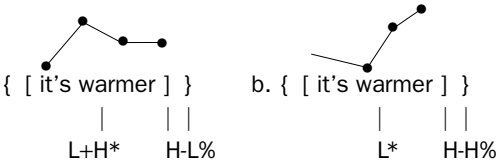
(12) PB DOWNSTEP: H $\rightarrow$!H / T + T . . . ___ . . . T% (Implementation)

(13)



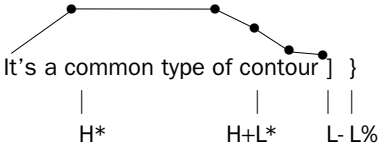
(14) PB UPSTEP: T% $\rightarrow$ raised T% / H- ___ (Implementation)

(15)



In addition to the two implementation rules given in (12) and (14), and the abstract downstep-trigger L of H*+L (see section 7.2.3), a final implementation rule was that L* of H+L* was realized as a downstepped H, H+L* therefore being equivalent to H+!H*. This pitch accent is used to describe the contour consisting of high-level pitch followed by a downstepped H* illustrated in (16).

(16)



7.2.8 ToBI

The somewhat abstract tonal grammar (11) was replaced with the practically oriented ToBI ('Tones and Break Indices') transcription system in 1992 (Silverman *et al.* 1992). ToBI combines a system for annotating prosodic boundary strength, essentially a five-point scale running from '0' for the boundary between a word and a cliticized form to '4' for 'end of sentence', with a user-friendly version of (11). Its creation was motivated by the need for large prosodically annotated text corpora and the difficulty that transcribers experienced when applying the Pierrehumbert–Beckman 1986 grammar. The main difference with Pierrehumbert and Beckman's (1986) analysis is that downstep is explicitly indicated in the transcription, in the spirit of Ladd (1983). The following specific changes were made.

(a) A downstepped H* is given as !H*. For instance, {L+H* L+H* L-L%} is {L+H* L+!H* L-L%} in ToBI.
(b) A new pitch accent was introduced to describe high-level pitch followed by downstep, H+!H*. For instance, {H* H+L* L-L%} (cf. (16)) is {H* H+!H* L-L%} in ToBI. Since H+L* is no longer needed to function as H+!H*, it was removed from the symbol set.
(c) Because H*+L was in effect a downstep trigger, with no realization of L, it was removed from the symbol set. For instance, {H*+L H* L-L%} is {H* !H* L-L%} in ToBI.

The pitch accents that ToBI is left with, therefore, are H*, L*, L+H*, L*+H, and H+!H*, while the downstep symbol is used to create !H* and L+!H*. Since H- can be downstepped too, H*!H- L% can be used to describe the vocative chant. The boundary tones, and their phonetic interpretations, were carried over intact. The only change in convention here is that in ι-final position the phrase accent, T-, is given as part of a complex symbol together with the boundary tone, T%, as in L-H%. Earlier, the phrase accent was printed where a detectable change occurred in the F_0 track, as in (5) and (6). This follows from the fact that the phrase accent was reanalysed as a boundary tone of the Intermediate Phrase in (1986) (cf. ((13), (15), and (16)).

Meanwhile, similar systems have been developed for a number of other languages, like German-ToBI and Korean-ToBI. A good overview is Jun (2003), who renamed the original system as AE-ToBI to make it explicit that it is an analysis of American English. To qualify for the label 'ToBI', a system should include the ToBI software enabling the transcriber to enter annotations on a number of time-aligned tiers. In addition to the tone tier and the Break Index tier, there is a tier for an orthographic transcription, one for disfluencies, and one for comments. Entries on all tiers are anchored to timing-points in the waveform, to be indicated by the transcriber.

Chapter 15 presents an alternative analysis of English intonation; in section 15.8, I will briefly evaluate the analysis represented by Pierrehumbert and Beckman (1986) and AE-ToBI.

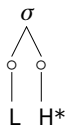
7.3 Developments since 1986

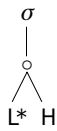
This section briefly discusses the developments since 1986, and is structured on the basis of tone types: pitch accents, boundary tones, plus the notion of 'secondary association', and phrase accents.

7.3.1 Pitch accents

By definition, accented syllables are associated with a pitch accent. Functionally, pitch accents may be lexical, as in Japanese or Swedish, or intonational, in which case they are frequently focus-marking, as in English. One of the tones, T*, associates with the TBU, any leading or trailing tone remaining unassociated. A focus-marking pitch accent typically increases the durations of segments in, and to some extent near, the accented syllable, as is the case in English and Dutch (Beckman and Edwards 1990; Cambier-Langeveld and Turk 1999; Cambier-Langeveld 2000), as it is for the functionally equivalent focus-marking tone of Swedish (Heldner and Strangert 2001) (accentual lengthening).

The tone whose target is closer to the rhyme of the accented syllable will typically be designated as the starred tone, but this decision may additionally be based on the more constant timing of its target relative to some segmental point in the accented syllable. Not all pitch accents appear to be timed in this asymmetrical fashion, however. Work on Standard Greek has shown that the low target of the pre-nuclear LH pitch accent occurs just before the stressed syllable and the high target at the CV boundary of the next, causing the rise to vary in duration depending on the number of consonants before and after the stressed vowel (Arvaniti, Ladd, and Mennen 2000). An implication is that it is no longer clear which is the starred tone, L or H. The pitch accent could be interpreted as a branching structure, which as a unit associates with the accented syllable, as the authors suggest. The internal structure of the bitonal pitch accent had earlier been discussed by Grice (1995b), who recognized a 'tone cluster' by the side of a 'tone contour', adopting a tonal root node between the syllable and the tones, as in Yip (1989) (see section 3.3.4). Grice claimed that in English, pitch accents with leading tones like L+H* are clusters (cf. (17a)), but that pitch accents with trailing tones like L*+H are contours (cf. (17b)). One of her arguments is that leading tones tend to be truncated when the accent is ι-initial, while ι-final trailing tones are not. To return to the Greek case, a structure like (17a) could be provided with a star for the whole complex to serve as the pre-nuclear LH (Arvaniti, Ladd, and Mennen 2000).

(17) a. 

b. 

Timing characteristics have been used by Frota (2002) to decide between an analysis of the European Portuguese focal and non-focal pitch accents as either bitonal H*+L and H+L* or as HL combinations of boundary tone and single-tone pitch accent. The evidence leads her to conclude that both are bitonal pitch accents, but that only the targets of H*+L are characterized by a constant interval.

Detailed phonetic studies of timings of targets have, on the one hand, revealed cross-linguistic and cross-varietal variation, and, on the other, a tendency for targets to be coupled to 'segmental landmarks', like the syllable offset, as in Mandarin lexical tones (Xu 1998), or the CV-boundary (Ladd, Faulkner, Faulkner, and Schepman 1999), while they can be sensitive even to the tenseness of the vowel (Ladd, Mennen, and Schepman 2000). Evidence for the influence of the length of the onset and the sonorant status of consonants is provided in Prieto, van Santen, and Hirschberg (1995), and Rietveld and Gussenhoven (1995), while the role of constraints like the speed of F_0 movements and the time difference between implementation and articulatory effect are discussed in Xu (2002).

The assumption that the targets of bitonal pitch accents are close together is not generally shared. The timing of trailing tones in English was made dependent on the distance to the next accent in Gussenhoven (1983b, 1988, 1999), Féry (1993), and Grabe (1998a). Part of the motivation was functional, as illustrated by (18a), after Gussenhoven (1983b), which has two occurrences of what intuitively appears to be the same neutral pitch accent. The target of the trailing L of the non-final H*L is timed rightmost, and bounded by the following associated tone. The trailing L of the final pitch accent is, however, constrained so as to occur immediately after the target of its T*. Because the timing of the trailing tone is context-dependent, the '+' is avoided, as it suggests that the tones are always realized close together. The representation of contour (18) in Pierrehumbert's 'on-ramp' theory is (18b), which has the pitch accents, H* and L+H*. A second argument for preferring (18a) over (18b) is that, unless the accents are close together, informal observations suggest that both the timing and the scaling of the low target before the second peak are imprecise; the target could be higher and earlier with no appreciable perceptual difference, suggesting that L's rightward drift is variable. This is not what is suggested by the pitch accent L+H*, however, where L's target would be expected to be located at a fixed distance from that of H*. A third argument is that the right-moving trailing tone, or 'displaced' tone, to use Grabe's (1998a) term, also occurs in non-final L*H and H*LH (Gussenhoven 1983b) and that therefore the description generalizes across pitch accents. The rightward displacement was termed a 'partial linking' in Gussenhoven (1983b), where 'linking' referred to the coherence of the two pitch accents,

and was seen as a step towards 'complete linking', the deletion of the trailing tone.

(18)

a. { Toronto is the capital of Ontario }

L_ι H*L H*L L_ι

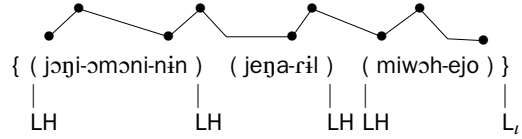
b. H* L+H* L-L%

7.3.2 Boundary tones

The reality of prosodic constituents is apparent from a number of phenomena. The context for segmental processes like assimilation is often defined by the boundaries of specific prosodic constituents (e.g. Nespor and Vogel 1986). Second, their boundaries reveal themselves through lengthening at the end (Wightman, Shattuck-Hufnagel, Ostendorf, and Price 1992; Gussenhoven and Rietveld 1992) and segmental strengthening at the beginning (Fougeron and Keating 1997; Fougeron 2001; Cho and Keating 2001). For instance, in *Tiptoe through the tulips,* the υ-initial [t] in *tip-* will be longer and have a more extensive articulatory contact than the ω-initial [t] in *tu-*, which in its turn will be longer and stronger than the F-initial [t] of *-toe.* Abstracting away from the presence of pitch accents and segmental effects on duration, *-toe,* the last syllable of the ϕ, will be longer than *tu-,* the first syllable of a ω, due to final lengthening. And third, syntactic movement rules may be sensitive to the size of the constituents they manipulate (Inkelas 1989; Inkelas and Zec 1990).

In intonation, the prosodic structure plays two roles. First, it may codetermine the context of phonological or phonetic rules (e.g. downstep, which is always confined to some prosodic constituent), including those responsible for the distribution of pitch accents (see section 7.4). Second, to return to the topic of this section, prosodic constituents may be marked by boundary tones. One, sometimes two, higher prosodic constituents may come with boundary tones initially and/or finally. These are now reported so frequently that they may well be universal. Some languages have *only* boundary tones, like Unangan, which has L_ϕ at the beginning and H_ϕ at the end of every ϕ. The only function of intonational structure in this language is thus to signal phrasing (Taff 1997). Boundary tones may be complex, i.e. consist of a tone sequence. Jun (1993) gives an analysis of Seoul Korean where every α has a string LHLH, which is reduced to just an initial LH if the α has only three syllables (Jun 1998). In the final α, any second occurrence of LH is pre-empted by one of the ι-final boundary tones H_ι, L_ι, $L_\iota H_\iota$ or $H_\iota L_\iota$, which express different intonational meanings. An example is (19). Formally and functionally, Korean is thus more complex than Unangan but, like Unangan, lacks pitch accents.

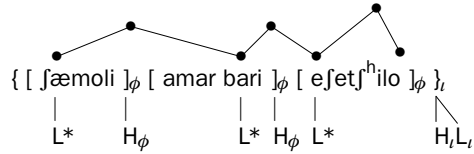
(19)



Youngi's mother-TOP Younga-ACC hate-ENDING
'Youngi's mother hates Younga'

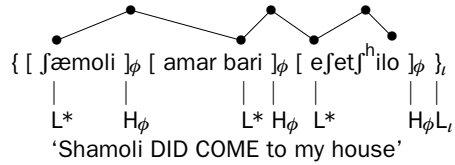
Bengali combines pitch accents and boundary tones. It closes the ι either by L_ι, or by one of the boundary complexes $L_\iota H_\iota$, for continuation, and $H_\iota L_\iota$, for yes–no questions. This is illustrated in (20), where L^* marks the accented syllables (see also section 4.2.2, figure 4.3). Complex $H_\iota L_\iota$ contrasts with $H_\phi L_\iota$, which contour is used to mark narrow focus with declarative intonation and is realized with an earlier and lower peak, as shown in (21) (Hayes and Lahiri 1991a). The F_0 peak signalling the question is always ι-final and high. Moreover, H_ϕ can move to a non-final ϕ, if the focus ends early, as shown in (22).

(20)



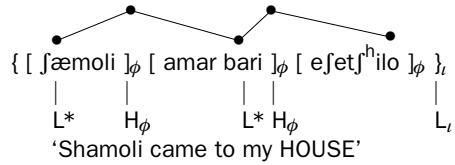
Shamoli my house come-PAST
'Had Shamoli come to my house?'

(21)



'Shamoli DID COME to my house'

(22)

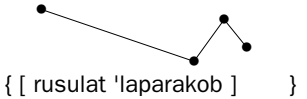


'Shamoli came to my HOUSE'

These examples also illustrate that more than one prosodic constituent may come with boundary tones, since in addition to H_ι, Bengali has H_ϕ. Evidence that H_ϕ is a boundary tone, rather than the leading tone of a possible $H+L^*$ pitch accent, is given by the pronunciation of *amar* 'my' in (20)–(22). The L^* goes to the first syllable of the first lexical word in the ϕ, and thus skips the function word *amar*. Because *amar* is pronounced on a downward slope, H_ϕ cannot be interpreted as the leading tone of the following pitch accent, since this analysis would predict that *-mar* has high pitch. Neither do Bengali sentences begin with high pitch.

Maori equally has a $L^*H_\phi L_\iota$ interrogative contour, but the middle tone here is a boundary tone for ϕ (de Lacy draft). Since Maori aligns the right edge of the focus constituent with ϕ and disallows pitch accents after the focus, a contour with an early peak is produced which in Bengali serves as the *declarative* early focus contour. For Georgian, Bush (1999) reports a complex boundary $L_\phi H_\phi$, as in (23), which is part of a question intonation contour that begins with H_ι. As Bush points out, the L-tone is not a pitch accent, as it spurns the stressed syllable in '*laparakob*. Neither is it a sequence of $L_\phi H_\iota$; one of his arguments is that the end of ι may contribute a L_ι in polite speech which could not be abstracted out of the contour if it was combined with the preceding H-tone.

(23)



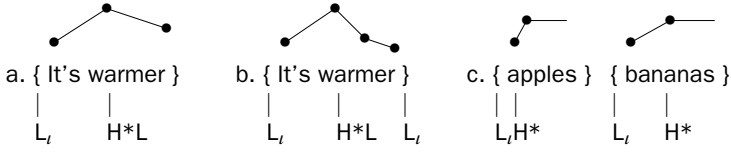
Russian you-speak

'Do you speak Russian?'

The Dutch dialect of Venlo has boundary tones of ι and υ (Gussenhoven and van der Vliet 1999): all ιs end in H_ι or L_ι, while the utterance can have an additional H_υ, leading to four utterance-final contours, H_ι, L_ι, $L_\iota H_\upsilon$, and $H_\iota H_\upsilon$.

Finally, as will be clear from these examples, boundary tones can be optional. Pierrehumbert's decision to have final obligatory boundary tones at two ranks may have led to a tendency to assume obligatory boundary tones in other languages. For English, this assumption is intimately connected with the assumption of two intonationally defined phrases.[1] In chapter 15, an analysis with optional final boundary tones will be presented. To give some examples of this analysis, (24a) gives a half-completed fall, which is contrasted with the fall to low in (24b), while (24c) shows a sequence of high levels, one type of listing intonation. In Pierrehumbert (1980), these three contours are transcribed H*+L H-L%, H* L-L%, and H* H-L%, respectively.

(24)

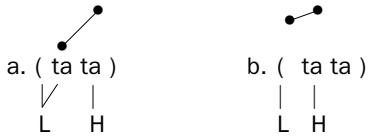


Likewise, Grabe (1998a) argues that an analysis of both German and English with an optional, single-rank boundary tone is to be preferred to a system with two constituents and obligatory boundary tones. Both the transcription system developed for Dutch, *Transcription of Dutch Intonation* (ToDI) (Gussenhoven, Terken, and Rietveld 1999) and the partly phonetic transcription system developed for varieties of British English, IViE (Grabe 2001) have optional boundary tones at a single rank only.

7.3.3 Secondary association

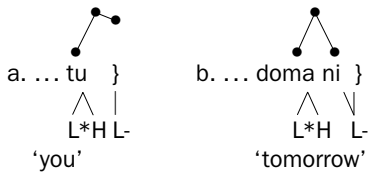
Pierrehumbert and Beckman (1988) introduced the concept of a boundary tone associating with a TBU. Their case will be dealt with fully in chapter 10, but the general idea is easily explained. The phrase in (25a) begins with a boundary L-tone and a toneless ('free') initial syllable, the second syllable being associated with H. In this situation, a low target occurs on the first syllable. In (25b), by contrast, the first syllable is associated with H, leaving no time for a low target: the pitch may rise a little towards the target for H at the beginning of the vowel, but nothing like the wide rise of (25a) will occur. Pierrehumbert and Beckman (1988) described this difference phonologically by associating the L-tone in (25a) with the first syllable, as shown, and assumed that the boundary tone with such a 'secondary association' (the 'first' association being with the constituent node) is pronounced fully low, as opposed to the L in (25b), which can only have a mid target.

(25)



The notion of secondary association was applied by Grice (1995a) to account for the variation between the slight fall on *ι*-final accented syllables in Palermo Italian question intonation, as in (26a), and the deep fall observed when the accented syllable is non-final, as in (26b).

(26)

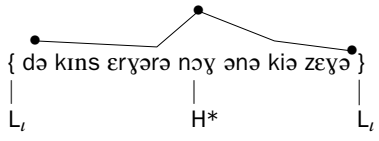


There is no necessary implication, however, that a tonal representation like (26a) inevitably leads to a phonetic implementation whereby the final L is not or is barely pronounced. Grønnum (1991) and Ladd (1996: 133) would describe a phonetic contour like (26a) as a case of truncation, and distinguish it from 'compression', i.e. a contour which reaches low pitch. Languages or language varieties may differ in that one is compressing and another truncating, as shown by Grabe (1998b) for RP and northern Standard German, respectively (see section 1.4.4), and by Grabe, Post, Nolan, and Farrar (2000) for Cambridge English and Leeds English, respectively.

While the difference between compression and truncation might arguably be accounted for by language-specific implementation rules, confirmation of the moraic association of boundary tones was provided by data from Venlo Dutch.

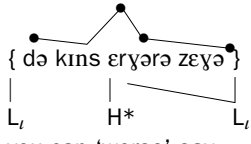
This dialect has a privative tone contrast on stressed syllables with two sonorant moras. There are thus three prosodic types of stressed syllable: those with one sonorant mora (27a); those with two but no tone (27b) (also known as Accent 1); and those with two and H on the second mora (27c) (also known as Accent 2). A focus-marking H* associates with the first mora of the stressed syllable of a focused word, and a declarative L_ι closes the ι. As shown in (27b), the fall for Accent 1 is completed inside the Accent-1 syllable, which is explained by the secondary association of L_ι with the sonorant mora in the accented syllable, a TBU which requires tone. Neither in (27a) nor in (27c) is such a 'free' mora available, and as a result the falls in the latter two contours are slower (Gussenhoven and van der Vliet 1999). Measurements for the contrast between (27b,c) were given in Peter van der Vliet's Master's thesis, while Gussenhoven (2000a) gives data for all three conditions in the related dialect of Roermond. When H_ι occurs instead of L_ι in situations like (27b), there are clearly two targets, forming a high level stretch.

(27) a.



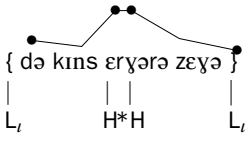
you can 'worse' AGAIN one time say

b.



you can 'worse' say

c.



you can 'irritate' say

Not only boundary tones acquire secondary associations. Chapter 11, section 11.2.2 will deal with the case of the trailing L of the Swedish pitch accent H*+L, which has been described as associating with a stressed syllable some distance away from its H* in compounds.

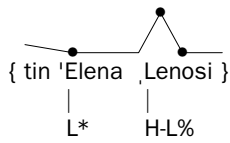
7.3.4 Phrase accent

The term 'phrase accent' has a chequered history. Pierrehumbert (1980) applied it to the internal boundary tone T-, equating it with Bruce's (1977) 'sentence accent', the focal H of Stockholm Swedish. What these tones have in common is

that they occur between the final boundary tone and the last T*(T), an intonational pitch accent in English and a lexical pitch accent in Swedish. Functionally, the Swedish 'sentence accent' is equivalent to the intonational pitch accents of English. In Beckman and Pierrehumbert (1986), the 'phrase accent' was reanalysed as a boundary tone of the ip, as noted above, allowing for an analysis in which lower ranking ips end in T- and higher-ranking ιs end in T-T%. Grice, Ladd, and Arvaniti (2000) narrowed the meaning down to a boundary tone with secondary association. This analysis explains why the English phrase accent (T-) is not in fact pronounced at the boundary, but usually immediately after the last pitch accent, and they therefore claim that T- associates with a stressed syllable inside the ip. They also suggest that the much-discussed issue of the difference between 'fall-rise' and 'fall-plus-rise' in descriptions of British English (cf. Cruttenden 1997) can be explained in terms of 'phrase accents'.

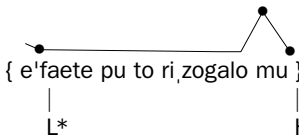
While the dependence of T- on stress may be less obvious in German or English, Grice *et al.*'s comparison between Standard Greek and Cypriot Greek makes the difference between an edge-seeking tone and a stress-seeking 'phrase accent' clear. Both languages have a yes/no question contour, analysed as L* H-L%. In Cypriot Greek, the H- remains close to the ι-boundary, with the peak usually falling in the last syllable, while in Standard Greek its target falls in the rightmost stressed syllable. In (28) for instance, H- associates with the unaccented, but main stressed syllable *Le-* in Standard Greek, while L% possibly does too, or else is timed after H-. (In (28), I place the L-target where it would appear to occur in the F_0 track.) The Cypriot Greek case is like that of Hungarian (Ladd 1983c; Gósy and Terken 1994; Varga 2002) and Bengali (Hayes and Lahiri 1991a), and shown in (29), where the rightmost stress is [zo], but the F_0 peak is on the unstressed final [mu].

(28)



the Elena Lenosi
'[Are you asking about] ELENA Lenosi?'

(29)

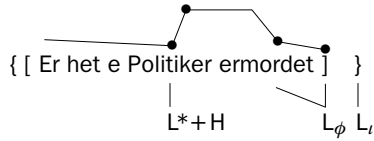


eat.2PL.PAST of the rice-pudding mine
'Have you eaten of my rice pudding?'

A third possibility is for the H-tone to have two high targets forming a plateau on post-focal stretches of speech, as occurs in Dutch tonal dialects (Gussenhoven and van der Vliet 1999) and Transylvanian Romanian (Grice, Ladd, and Arvaniti 2000). In Bern Swiss German, a similar plateau is due to H of the L*+H pitch

accent, which in broad focus continues until the last stressed syllable of the phrase, where a L_υ seeks a secondary association (Fitzpatrick-Cole 1999). This is shown in (30).

(30)



7.4 Rhythmic adjustments of pitch-accent distribution

In addition to contributing boundary tones, prosodic constituents may determine the rhythmic distribution of pitch accents. Generally, adjacency of prominent syllables is disfavoured (Alternation Principle, Selkirk 1986). This is as true for (accented and unaccented) stressed syllables within words as it is for adjacent pitch accented syllables, whether occurring within the same word or in different words. Many languages manage to live with such 'clashes'. English, for instance, allows adjacency of stressed syllables in underived words, as illustrated by *can'teen, 'syn,tax*, and, in a rare pattern of (x)(x .), *'cu,cumber*. Other languages do not. For instance, strong syllables never clash in the simplex word in Dutch (Gussenhoven 1993).

Clashes of accented syllables are, however, avoided in English, as indeed they are in French and Dutch: English *ponTOON*, but *PONtoon BRIDGE*, French *beauCOUP*, but *BEAUcoup PLUS* 'much more' and Dutch *bijGAAND*, but *BIJgaand BRIEFje* 'accompanying letter', but are tolerated in Portuguese, Italian, and Greek (Frota 1998; Farnetani and Kori 1983; Arvaniti 1994). Greek [po'li 'liɣa] 'too small', for instance, will not be pronounced *['poli 'liɣa] (Arvaniti 1994).[2]

There has been some confusion over the phonological nature of clash resolution in English ('stress shift'). Liberman and Prince (1977) proposed an analysis in terms of a metrical grid, following Chomsky and Halle (1986) and others who worked with a 'segmental' feature [*n*stress]. As a result, the insight that English clash resolution concerned the distribution of pitch accents (Bolinger 1965; Bolinger 1986; Vanderslice and Ladefoged 1972) was lost. The renewed claim that English clash resolution amounts to arrangement of pitch accents, typically amounting to a deletion of the leftmost of the two clashing pitch accents (Gussenhoven 1986; Gussenhoven 1991a; Shattuck-Hufnagel 1989; Ladd and Monaghan 1987), was confirmed by acoustic measurements (Horne 1990; Vogel, Bunnell, and Hoskins 1995), and found support in the Boston University Radio News Corpus described in Ostendorf, Price, and Shattuck-Hufnagel (1995), Shattuck-Hufnagel, Ostendorf, and Ross (1979), and Shattuck-Hufnagel (1995). For instance, a word like *Japanese* is typically pronounced as in (31a) when it is the only accented word in the utterance. Depending on its position in the ϕ, either the first or the second accent will be deleted, giving (31a) and (31b), respectively. The latter is said to have undergone 'stress shift' in Liberman and Prince (1977).

Acoustic measurements show that clash resolution in French is likewise a matter of pitch accent distribution (Post 2000a; Post 2000b).

(31) a. It's JAPaNESE
b. It's GOOD JapaNESE
c. It's a JAPanese PRINT

Word prosodic structure is involved in the clash resolution of accents in English to the extent that a pitch accent cannot be relocated on an unstressed syllable (Liberman and Prince 1977). Compare (32a), which contains an adjective with two feet, with (32b), which consists of an unstressed syllable and a monosyllabic foot. Similarly, compare (32a) with (32c), from Dutch, which does not allow adjacent footheads, and in which *Chinees* has the word prosodic structure of English *obese.*

(32) a. a CHInese BOOK (cf. CHI-NESE) b. an oBESE PERson c. een ChiNEES BOEK 'a Chinese book'

The rhythmic redistribution of accents in French and English will be dealt with extensively in chapters 13 and 15.

7.5 Conclusion

On the one hand, this chapter has pulled together information about intonational phonology that already figured in earlier chapters, while on the other it preluded on the language descriptions in chapters 9–15. The next chapter explores the way these intonational structures can be described in Optimality Theory. It is possible to skip that chapter and still profitably study the language descriptions, since the optimality-theoretic element in those descriptions concerns certain selected aspects only.

Notes

1. In my description of standard Dutch, L% and H% were explicitly added in Gussenhoven (1991a). Earlier they had been understood as part of the phonetic realization of 'basic' H*L and L*H in nuclear position (Gussenhoven 1983a), where they contrasted with 'half-completed' versions of the same contours. Absence of L% and H% in these contours came to stand the half-completed realizations in a synthesis-by-rule programme (Gussenhoven and Rietveld 1992).
2. In Greek, the clash is phonetically resolved by lengthening of the vowel of the first syllable in the clash, and to a lesser extent, of the onset consonant of the second syllable. In Italian, onset consonant lengthening is phonological (Raddoppiomento Sintattico), while first vowels are phonetically lengthened (Farnetani and Kori 1983; Nespor and Vogel 1986).