Intonation systems across varieties of English

Martine Grice, James Sneed German, Paul Warren

Chapter 19
Intonation systems across varieties of English

Martine Grice, James Sneed German &Paul Warren

Revised version 25th August 2019 for preprint
In Gussenhoven, Carlos & Aoju Chen (scheduled to appear October 2020) The Oxford Handbook of Language Prosody, Oxford: OUP.

19.1 The role of English in intonation research

The mainstream standard varieties of English have played a major role in the development of models of intonation, with different traditions either side of the Atlantic. The British School emphasized auditory training, including the production and perception of representative intonation patterns - akin to training in the cardinal vowel system (Jones 1967) - and transcription using tonetic stress marks. These diacritics indicate the position of stressed syllables and pitch movements (falling, rising, level, falling-rising and rising-falling) across these and following syllables. The most important feature is the *nuclear contour*, the pitch movement over the stressed syllable of the most prominent word in the phrase[1] (nucleus) and any following syllables (tail). The British School has included didactic approaches (O'Connor and Arnold 1961), approaches focused on phonetic detail and broad coverage (Crystal 1969), and approaches considering the relationship with syntax and semantics/pragmatics (Halliday 1967; Tench 1996, amongst others).

Meanwhile, American Structuralism focused on phonological structure, separating stress-related intonation from intonation marking the edges of phrasal constituents (Pike 1945; Trager and Smith 1951). Together with the all-encompassing work by Bolinger (1958, 1986, 1989), who introduced the concept of pitch accent, and insights from work on Swedish (Bruce 1977), this set the stage for autosegmental-metrical (AM) approaches to English intonation (*inter alia* Liberman 1975; Leben 1976; Pierrehumbert 1980; Beckman and Pierrehumbert 1986; Gussenhoven 1983; Ladd 1983).

In AM theory, the separation of prominence-lending and edge-marking aspects of intonation is crucial: pitch accents are associated with stressed syllables, and edge tones with phrase edges. The AM equivalent to the British School nuclear tone is the combination of the last pitch accent and the following edge tones, referred to as the nuclear contour. In what follows, we couch our discussion in AM theory, currently the most widespread approach to intonation, paying special attention to the nuclear contour, which is often the focus of comparison in the papers we have consulted. Although using AM representations facilitates comparison, we nonetheless have to exercise caution, since AM models can differ from each other in ways that do not always reflect differences in the varieties they describe. Each model can be seen as a prism through which to observe a variety, and at the same time a model that is developed to describe a variety is likely to have been shaped by this variety. We return to this problem when discussing mainstream varieties below.

19.2 Scope of the chapter

This chapter is concerned with the structure and systematicity underlying intonation. A historical emphasis on standardized varieties from the British Isles and North America means that our descriptive understanding of those varieties is more comprehensive, and they have also played a central role in the development of theoretical models and frameworks. Therefore, we treat those varieties, and closely related varieties from Australia, New Zealand,

[1] Whether the nuclear syllable is always the most prominent in the phrase has been contested. Instead it has been defined positionally as the last content word or the last argument of the verb.

and South Africa, under the label 'Mainstream English Varieties' (henceforth MEVs). MEVs show relatively few differences in overall phonological organization, whereas more substantial variation arises in cases where MEVs were adopted and subsequently nativized by non-English-speaking populations. We therefore examine separately a selection of 'Contact English Varieties' (CEVs) that differ from MEVs in typologically interesting ways[2]. We explore the challenges posed by their diverging prosodic structures, in terms of both prominence and edge-marking, and observe that an account of the intonation of current-day 'Englishes' needs to cover a broad range of typological phenomena going well beyond what is present in the extensively researched mainstream varieties. This broad range in turn provides us with a chance to observe prosodic variation within one language that is usually only available to crosslinguistic studies.

19.3 Intonational systems of Mainstream English Varieties

MEVs have many intonational properties in common, both in terms of the distribution of tones throughout the utterance and in terms of their local and global tonal configurations. Hence, although we discuss Northern hemisphere and Southern hemisphere varieties separately below, this is mainly for convenience in the context of this handbook, which groups languages according to their geographical location.

19.3.1 Northern hemisphere

All MEVs have lexical word stress, in that specific syllables within a word are designated as prosodically privileged in the sense of Hyman (2014) and Ladd (2008a). Not only do such syllables receive acoustic and articulatory enhancement, but they are also possible locations for pitch accent realization. When present, pitch accents usually occur on the primary stressed syllable of a word, though in particular contexts, secondary stressed or even unstressed syllables can be promoted to bear accents. This can be for rhythmic reasons – *Chi'nese,* has an initial accent in *CHInese GARden* to avoid a clash of pitch accents on adjacent syllables – or in metalinguistic corrections, such as '*I met roSA not roSIE*'. Northern Hemisphere MEVs also share a considerable reduction of unstressed syllables leading to strong durational differences between stressed and unstressed syllables. This is not always the case in Southern Hemisphere MEVs (see §19.3.4).

Pitch accents are sparsely distributed in MEVs, and are more common on nouns than on verbs, and on content words than function words. Each intonation phrase (or intermediate phrase if the model includes one) requires at least one pitch accent – the nuclear pitch accent (with exceptions for subordinate phrases and tags, see Crystal 1975: 25; Firbas 1980; Ladd 2008a: 238, Gussenhoven 2004: 291). Prenuclear pitch accents are not usually obligatory in MEVs, but may serve information structural functions (e.g. topic marking). They may also be optionally inserted for rhythmic purposes, usually near the beginning of a phrase.

MEVs use the placement of nuclear pitch accents to mark the difference between new or contrastive information versus discourse-given information. The rules governing this relationship are complex and outside the scope of this chapter (see Wagner 2012 for a review). In broad terms, however, a focused constituent obligatorily contains a nuclear pitch accent, while discourse-given words following the focused constituent are unaccented. This contrasts with many contact varieties, for which accent placement does not appear to be used to mark information status or structure.

The intonation systems of MEVs have a rich paradigmatic range of pitch accents and edge tone combinations. The pitch accents proposed in a consensus system (Beckman et al. 2005)

[2] This distinction is closely similar to that between Inner and Outer Circle varieties (Kachru 1985). A separate terminology is used here because the present review is concerned primarily with similarities and differences in the synchronic structural aspects of nativized varieties, as opposed to their sociolinguistic contexts.

for describing Mainstream American English (MAE) are simple low (L*) and high (H*) tones, rises (L+H*), and scooped rises (L*+H), downstepped variants of each of these last three (!H*, L+!H*, L*+H!), and an early-peak falling accent (H+!H*). This system assumes two phrase types: the intonation phrase (IP) and the intermediate phrase (ip). The latter has either a high (H-), low (L-), or downstepped high-mid (!H-) edge tone at its right edge. IP-finally, the ip edge tone is followed by a high (H%) or low (L%) IP edge tone (Pierrehumbert 1980, Beckman and Pierrehumbert 1986). The co-occurrence of ip and IP edge tones leads to complex tone sequences at intonation phrase edges: A phrase-final nuclear syllable – if combined with a bitonal pitch accent – can carry up to four tones. These pitch accents and edge tones are illustrated in online ToBI (Tones and Break Indices) training materials (Veilleux, Shattuck-Hufnagel and Brugos 2006).

The edge tones of ips are also referred to as *phrase accents*. For some varieties it has been argued that they can be associated with a postfocal stressed syllable (Grice, Ladd and Arvaniti 2000)[3], lending them an accent-like quality. Although this does not promote the syllable sufficiently to allow it to bear a pitch accent (thus, it does not counteract deaccentuation), the syllable is rendered more prominent by virtue of bearing a tone.

This same inventory of pitch accents and edge tones could in principle be used to describe Southern Standard British English (SSBE) (Roach 1994; Ladd 2008a). However, there are a number of differences in the tonal inventories of AM models developed for SSBE, although they appear to reflect differences in the models themselves rather than differences in intonation (MAE and SSBE). For instance, some models include a pitch accent representing a nuclear fall H*+L (also written H*L, see Gussenhoven 1983, 2004 on British English; Grabe, Post and Nolan 2001), whereas others capture this configuration with a sequence of pitch accent and ip edge tone H* L-. AM models of MEVs also differ in their treatment of the movement (onglide) towards the starred tone. In MAE-ToBI it is captured with a leading tone (L+H*), whereas in some other models it is either non-distinctive (and therefore treated as phonetic detail, Grabe et al. 2001) or derived from an L tone from a previous accent (Gussenhoven 2004). Nonetheless if the pitch movement is falling (i.e. there is high pitch before the accented syllable), there is a general consensus across the different models that this should be captured with an early peak accent (H+H*).

The status of the onglide is certainly of theoretical import, but does not capture differences across the intonation of these varieties (see Grice 1995 for a discussion). It may be, for instance, that a more frequent use of falling pitch phrase-medially leads to the conclusion that there must be a H*+L pitch accent (Estebas Vilaplana 2003), whereas if falls tend to occur phrase finally, they might be more likely to be analysed as H* followed by a L- edge tone. The danger of comparing different intonational systems using models that are based on different assumptions, even if they are all in the AM tradition, is discussed in Ladd (2008b), who argues that this can lead to comparisons that are not about particular languages or varieties but about the models themselves. Crucially, differences in the analysis of specific patterns should not be downplayed, as they have consequences for the overall organization of a model, including, for example, the need for an ip-level of phrasing. Ladd (2008a) and Gussenhoven (2016), among others, provide valuable critiques of the MAE-based consensus model outlined above. However, they highlight that revisiting early theoretical choices of that model need not undermine the appropriateness of an AM approach to MEVs.

Although any combination of pitch accents and edge tones is in principle allowed, certain preferred combinations occur more frequently than others (Dainora 2006). Likewise, although individual pitch accents and edge tones have been assigned pragmatic functions (Pierrehumbert and Hirschberg 1990; Bartels 1999), it is often combinations of nuclear pitch

[3] This is particularly common in fall-plus-rise contours, referred to as a compound fall-plus-rise tune in the British School, e.g. 'My \mother came from /Sheffield', where Sheffield is given in the discourse context (O'Connor and Arnold 1961:84).

accents and edge tones (i.e., nuclear contours) that are referred to when discussing the meaning of intonation (see e.g. Crystal 1969; Brazil et al. 1980; Tench 1996; Cruttenden 1986). For instance, rise-fall and (rise-)fall-rise nuclear contours are said to convey meanings such as unexpectedness or disbelief respectively (see §19.5 for discussion of rises). Differences between North American and British standard varieties can often be expressed in terms of which nuclear contour is preferred in which pragmatic context (Hirst 1998), rather than differences in the inventory. However, differences in usage can lead to misunderstandings: For instance, a request using H* L-H% is routinely used in SSBE but can sound condescending in MAE (Ladd 2008a), where a simple rise is preferred.

19.3.2 Non-mainstream varieties of American English

The limited research on the intonation of regional or ethnic varieties of AE suggests that varietal differences are relatively minor. Comparing rising pitch accents in speakers from Minnesota and Southern California, for example, Arvaniti and Garding (2007) found that the former group has later alignment of tonal targets and likely lacks a distinction between what are taken to be two distinct accents in the latter variety, H* and L+H*. In a comparison of Southern and Midland US varieties, Clopper and Smiljanić (2011) found distributional differences, namely in the relative frequencies of certain edge tone categories, which varied also across gender and texts.

African American English (AAE) has been relatively well-studied (see Thomas (2007; 2015) for a review). One widely observed feature is that certain words have initial stress where MAE has non-initial stress (e.g., *ˈpo.lice*, *ˈho.tel*; Fasold and Wolfram 1970, *inter alia*). This is noteworthy since it affects tonal alignment. A few studies report final falling or level contours in AAE for polar questions (Loman 1975; Tarone 1973; Jun and Foreman 1996; Green 2002), whereas MAE speakers typically use a high rising contour (L* H-H%). Jun and Foreman (1996) also report post-focal pitch accents (i.e. no deaccenting) with early focus. Other differences involve f0 scaling, including a greater overall range (Tarone 1973; Hudson and Holbrook 1981; 1982; Jun and Foreman 1996), and a lower degree of declination (Wolfram and Thomas 2002; Cole et al. 2008).

Research on Chicano English (CE) in Hispanic communities remains entirely descriptive, and characteristics vary by region. Metcalf (1979) provides a comprehensive survey of CE studies spanning 20 years, identifying five features as typical: (i) a tendency for noun compounds to have primary stress on the second constituent rather than the first (e.g., *baby ˈsitter*; Metcalf 1979; Penfield 1984), (ii) a less pronounced utterance-final fall for declaratives and wh-interrogatives (Castro-Gingras 1974; Metcalf 1979), (iii) greater use of non-final rising pitch prominences (Castro-Gingras 1974; Thomas and Ericson 2007), (iv) the use of final falling contours for polar questions (Castro-Gingras 1974; Fought 2003), and (v) the use of emphatic ‘rising glides’ with a peak very late in the stressed syllable (Penfield 1984; Penfield and Ornstein-Galicia 1985). Fought (2002) also notes additional lengthening of stressed syllables at the beginnings and ends of utterances.

Burdin (2016) presents a comprehensive examination of English intonation amongst Jewish and non-Jewish Americans in Ohio. While the differences mainly concern category usage (e.g., Jewish speakers use more rising pitch accents in listing contexts and more level contours in narratives), her findings suggest that contact with Yiddish has contributed to a more ‘regular alternation between high and low tones within a prosodic phrase’ (p. 12), or *macro-rhythm* (Jun 2014), in Jewish English.

In general, existing research on non-standard AE varieties is highly descriptive, rarely concerned with phonological structure, and mostly from older studies; this area of research clearly needs updating.

19.3.3 Non-mainstream varieties in the UK and Ireland

Across British varieties there is considerable regional and individual variation in intonational inventories (Cruttenden 1997; Grabe et al. 2000), as well as differences in preferred patterns in given pragmatic contexts. One striking aspect of many urban Northern varieties (e.g. in Birmingham, Liverpool, Glasgow, Belfast and Tyneside, which Cruttenden (1994) dubbed Urban Northern British (UNB)), is the routine use of rising contours in statements. Not only their form but also their usage as the default contour distinguishes them from uptalk (see §19.5). Cruttenden (1997) proposes four types of 'rise': a simple rise (preferred in Glasgow), two rises with plateaus (rise-plateau and rise-plateau-slump, both common in Birmingham, Liverpool, Belfast and Tyneside), and a rise-fall, found less commonly in several UNBs, as well as in Welsh English. An alternative AM transcription system was developed specifically to capture this regional variation (IViE, Grabe et al. 2001), transcribing these contours as L*+H (Grabe 2002, 2004) combined with different edge tone sequences.

To capture the difference between a final rise, a level and a slump, IViE incorporates a null boundary tone (Ø%) in addition to H% and L% (the latter taken to be low, not upstepped as in MAE ToBI). However, in a detailed description of Manchester English, Cruttenden (2001) argues that this addition is insufficient to capture the range of distinctive contours both here and in UNB varieties. Distinctions need to be made between slumps and slumps plus a mid-level stretch, and between simple falls and falls plus a low-level stretch. Cruttenden argues for a feature involving sustention, rather like spreading, which will be shown to be useful for contact varieties (section 19.4).

Irish English, referred to as Hiberno English (HE), is defined by a set of unofficial localized linguistic standards, setting it apart from British varieties. The intonation patterns in HE varieties can be mapped onto corresponding regional Irish varieties (Dalton and Ní Chasaide 2007; Dorn and Ní Chasaide 2016). There is a clear difference in the intonation of the South and the North of Ireland. According to Kalaldeh et al. (2009), declarative statements in Dublin and Drogheda English tend to have prenuclear and nuclear falling contours, analysed as bitonal H*L pitch accents, whereas across the geographical North of Ireland, e.g. in Donegal and in Belfast, they have prenuclear and nuclear rises (analysed as L*H). O'Reilly et al. (2010) report post-focal deaccenting in Donegal English. Interestingly, early focus in this variety, leads to a gradual fall from the focal H tone over the post-focal domain, reaching a low pitch phrase-finally, rather than the characteristic final rise of broad focus or late focus statements.

19.3.4 Southern hemisphere mainstream varieties

Many of the properties of Northern hemisphere varieties apply to Southern hemisphere varieties too. Analyses of Australian English (AusE) assume the same inventory of pitch accents, phrase accents and boundary tones as the AM analyses of Mainstream American English (see for instance Cox and Fletcher (2017)), although earlier analyses (e.g., Fletcher et al. 2002) included H*+L and its downstepped variant, !H*+L, which are absent from MAE_ToBI.

Rising tunes have been widely studied in AusE (see also §19.5). Fletcher et al. (2002) found that low-onset high rises (L* H-H%) and expanded-range fall-rises (H*+L H-H%) are more frequent and have greater pitch movements than low-range rises (L* L-H%) and low-range fall-rises (H* L-H%). There are higher top-line pitch values and more high-onset rises (H* H-H%) in instructions and information requests than in acknowledgements. For participants issuing instructions, expanded-range rises were more likely at the ends of turns than turn-medially.

An analysis of newsreader speech showed that New Zealand English (NZE) has comparatively fewer level nuclei and more complex nuclei (the latter potentially expressing

greater emotional involvement) than British English (Warren and Daly 2005). NZE tends to have more dynamic intonation patterns with a higher rate-of-change in pitch (Warren and Daly 2005) than British English, though this varies by region within New Zealand. Vermillion (2006) found that NZE is characterized by higher pitch, with higher boundary tones (H%) but smaller pitch falls between adjacent H* accents.

Comparisons of AusE and NZE have chiefly focused on rising intonation patterns, particularly uptalk (Fletcher et al. 2005; Warren and Fletcher 2016a, 2016b). One point of comparison has been how the two varieties distinguish uptalk rises on declaratives from yes/no question rises, with more dramatic rises for statements than for questions (see §19.5) and more fall-rise patterns for uptalk in AusE.

Moving finally to South Africa, given that English is the most commonly spoken language in official and commercial public life and is the country's lingua franca, it is surprising how little has been written about the intonation of this variety. Indeed, reference descriptions of South African English (e.g., Bowerman, 2008; Lass, 2002) make no mention of prosodic features. Uptalk rises on declaratives have been reported for White South African English (WSAfE), and as in NZE, these tend to have a later rise onset than question rises (Dorrington, 2010; Dorrington and Bekker 2010). For Black South African English §19.4.8.

19.4 English intonation in contact

This section considers varieties arising from L2 use of English, with subsequent nativization. Contact with typologically diverse languages has resulted in intonation systems that differ, sometimes dramatically, from those of MEVs.

19.4.1 Hong Kong English

Hong Kong English (HKE), or Cantonese English, refers to a variety spoken by L1 speakers of Cantonese, either as an L2 or as a nativized co-L1 in contexts like Hong Kong, where both languages are spoken side-by-side. Luke (2000) provides the first account of HKE intonation in terms of tone assignment that is sensitive to lexically stressed syllables. Word stress patterns of SSBE are interpreted in terms of three Cantonese level tones, High (H), Mid (M), and Low (L). The resemblance between Cantonese level tones and the pitch patterns of HKE syllables is indeed striking, and most authors assume that the tone inventory of HKE at least partly originates from Cantonese (though see Wee (2016) for a discussion). Luke (2000) proposes the rules in (1) for tone assignment in HKE, where 'stress' corresponds to all primary stressed syllables, and variably to secondary stressed syllables, in BrE. Other patterns result from concatenations of the word-level patterns, except that certain classes of function words (possessors, modals, and monosyllabic prepositions) are realized as M (Wee, 2016).

(1) a. Stressed syllables are realized as H.
b. Unstressed syllables before the first stressed syllable in a word are realized as M.
c. H spreads rightward

Luke (2000) assumes that all syllables have level pitch, though later studies show that the final syllable may be realized as a fall, resulting from a declarative L% boundary (Wee, 2008; Cheung, 2009). As Figures 19.1a and 19.1b show, the final syllable is realized as HL if it is stressed, and as L otherwise. For sentences that end with more than one unstressed syllable in sequence, an interpolated fall stretches from the H of the last stressed syllable (Figure 19.1c) (Wee 2016).

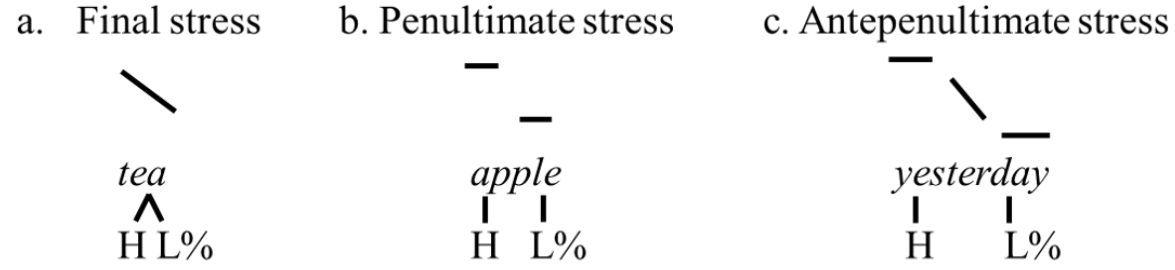

Figure 19.1. Tonal representations and stylized f0 contours for three stress patterns in a declarative context (adapted from Wee (2016: ex. 15d)).

Luke (2000) and Gussenhoven (2017, 2019) also observe a sustained, high-level final pattern for final stress utterances. Gussenhoven attributes this to an absence of boundary tone (Ø%), which contrasts with L% and conveys a non-emphatic declarative meaning. Wee (2016), however, argues that this is a phrase-medial pattern (even when pronounced in isolation), ruling out a boundary tone.

The right edge of polar interrogatives is high rising H H% if the last syllable is stressed (Figure 19.2a) and a falling-rising HLH% from the last stressed syllable otherwise (Figure 19.2b, 19.2c). According to Gussenhoven (2017, 2019), pre-final L in the latter case arises from an HL assigned to the last stressed syllable (as opposed to H for earlier stresses), while Wee (2016) assumes that LH% is a compound boundary tone.

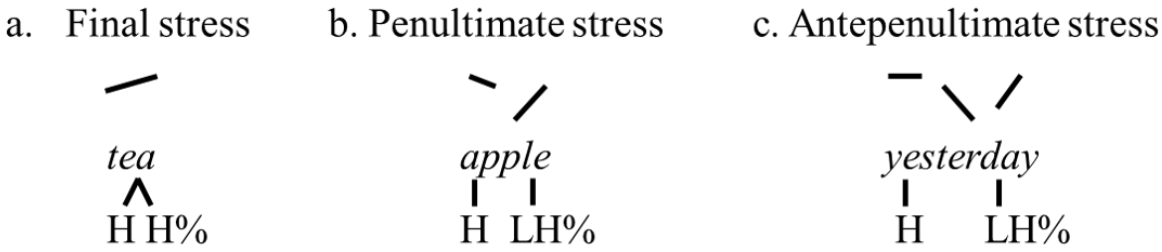

Figure 19.2. Tonal representations and stylized f0 contours for three stress patterns in a polar interrogative context (adapted from Wee (2016: ex. 15g)).

Existing accounts mostly agree on the surface phonological representations in HKE, and that deriving those representations begins with an assignment of H to lexically privileged syllables corresponding to stress in MEVs. They also agree that H spreads rightward within a word unless the word is IP-final, and M is assigned to unstressed syllables before the first stress in a word, as well as to certain functional categories.

19.4.2 West African Englishes (Nigeria and Ghana)

Nigerian English (NigE) intonation has been described as involving predominantly level tones (Wells 1982; Jowitt 1991; Gut 2005). This is not surprising given its emergence in contact with level tone languages such as Hausa, Igbo, and Yoruba. Gut (2005) proposes that tones are assigned to syllables based on word stress and grammatical category: Lexical words take H on all syllables except the first, which is L if unstressed and H otherwise. Certain functional categories (e.g., personal pronouns) also take H, while articles, prepositions, and conjunctions take L. Downtrending is widely observed (Jowitt 2000; Gut 2005). A perception study (Gussenhoven and Udofot 2010) suggests that this results from obligatory downstep between subsequent H tones, triggered by an L associated to the left edge of lexical words (in words with initial stress, L is floating). A further production study (Gussenhoven 2017)

indicated that, contra Gut (2005), H is assigned only to syllables that have primary or secondary stress in SSBE, and that intervening syllables have high f0 due to interpolation.

Word-level tone assignment appears obligatory, and cannot be modified for the expression of contrast or information status (Jibril 1986; Jowitt 1991; Gut 2005). This may explain the impression that NigE has nuclear pitch on the final word. For declarative utterances, the final syllable is falling, which Gussenhoven (2017) attributes to L%. Polar interrogatives generally end with a high or rising tune (Eka 1985; Jowitt 2000), suggesting that L% alternates with H%.

Ghanaian English (GhE) is very similar to NigE with a few notable exceptions. Gussenhoven (2017), building on observations by Criper (1971) and Criper-Friedman (1990), proposes that H is assigned only to syllables corresponding to *primary* stresses in SSBE. For lexical words and certain functional categories, a word-initial L spreads to all syllables before the first H, while H spreads rightward from that syllable. IP-finally, word-final unstressed syllables are low or falling, suggesting that H is blocked from spreading rightward in that position. As in NigE, most function words are assigned L. Downstep is triggered across subsequent LH-bearing words, while pitch reset occurs across IP boundaries. Polar interrogatives are not usually marked with a boundary tone, ending instead with the pitch level of the last H. Occasionally, the final syllable of a polar interrogative is low-rising; thus when H% occurs, a polar tone rule changes the final H to L.

19.4.3 Singapore English

Singapore English (SgE) is a fully nativized L1 variety. Virtually all Singaporeans speak SgE by early school age, and it is the dominant home language for one third of the population (Leimgruber 2013). SgE has been influenced by contact with a wide range of language varieties, and even highly standardized uses of SgE differ markedly from MEVs in terms of prosody (Tay and Gupta 1983).

SgE intonation has been described as a series of 'rising melodies', ending in a final rise-fall (Deterding 1994). The domain of rises is typically a single content word, along with preceding function words (Chong 2013), though in certain syntactic contexts, a rise may span two content words (Chong and German 2015, 2017). As Figure 19.3 illustrates, the utterance-initial rise has a large pitch excursion, while subsequent rises have smaller ranges (Deterding 1994; Low and Brown 2005). The repeating rises likely correspond to units of phrasing, as they reflect the primary domain of tune assignment and serve a grouping function.

Chong (2013) proposes an AM model based on the Accentual Phrase (AP). The AP is marked at left and right edges by phrasal L and H tones respectively, as well as by an optional L* pitch accent aligned with the primary stressed syllable of the content word. Chong's phrasing analysis is further supported by the fact that these rises are associated with extra lengthening of the word-final syllable (Chong and German 2017). Chong's (2013) model also includes two further levels of phrasing, ip and IP. The ip is marked at the right edge by H- (IP-medially) or L- (IP-finally) and accounts for pitch reset in adjacent APs. The IP is marked at the right edge by L% for declaratives and by H% for polar interrogatives.

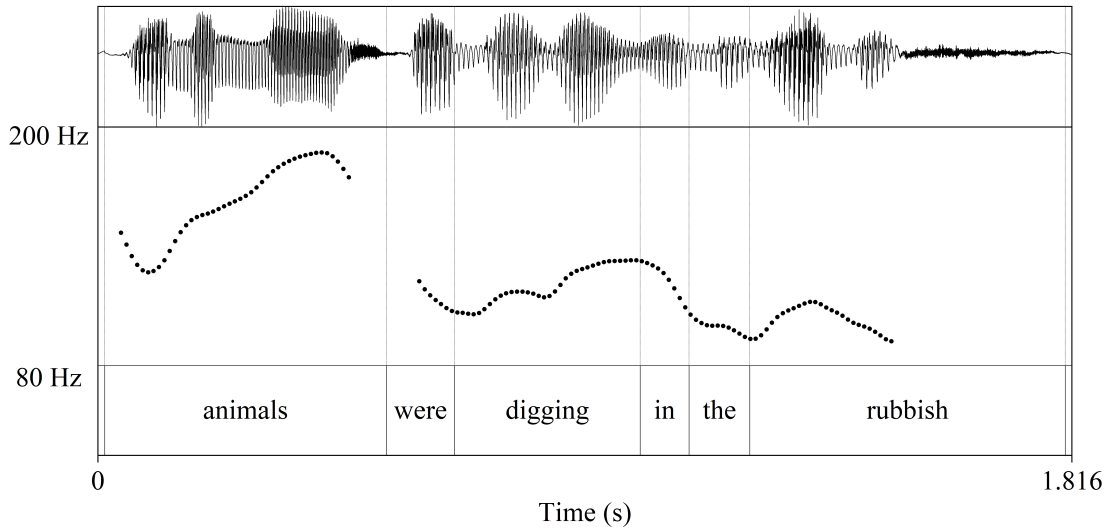

Figure 19.3. Waveform, spectrogram, and f0 track for a sentence of read speech in SgE (D'Imperio and German 2015: Fig. 1b).

F0 peaks are aligned close to the end of the word in non-final APs. Thus, if contrastive word-level prominence is present in SgE, the link to f0 differs from that in MEVs. SgE lacks marked f0 changes between stressed and unstressed syllables, as well as any step-down in f0 from stressed to post-stress syllables within a word (Low and Grabe 1999). The proposed stress-sensitivity of L* was not corroborated by Chong and German (2015), who found no effect of stress on the alignment of low or high targets in initial APs. Instead, the contour shape was similar across stress patterns, while words with initial stress were produced with a globally higher f0. The alignment of f0 peaks in IP non-initial positions needs systematic investigation.

Wee (2008) and Ng (2009, 2011) propose that three level tones (L, M, H) are assigned to syllables at the phonological word level. H is assigned to all final syllables, L to initial unstressed syllables, and M elsewhere either as the default tone or through spreading. Some syllables may end up toneless in which case interpolation applies. Further quantitative studies are needed to establish which aspects of SgE contours are better accounted for by level tones assigned to syllables versus phrase-level tones.

Existing research concentrates on ethnically Chinese speakers and read speech (though see Ng (2009)). Tan (2002, 2010), however, describes tune differences across three ethnic groups of Singapore (Chinese, Malay, and Indian), which reflect characteristics of the respective 'mother tongue' languages. More research is needed to establish which aspects of SgE intonation are shared across communities and registers.

19.4.4 Indian English

Indian English (IE) is widely considered to be a fully nativized variety (Kachru 1983, i.a.). Although there is a high degree of variation across speakers, IE intonation appears to conform to a single basic phonological organization similar to that of MEVs, while varietal differences concern mainly word-level stress, tonal inventories, or meaning.

Word-level stress in IE is largely rule-governed rather than lexically specified, though the specific rules differ by L1 background, region, and individual (Wiltshire and Moon 2003; Pandey 2015; Fuchs and Maxwell 2015). For instance, stress assignment is quantity-sensitive for Hindi or Punjabi speakers (Pandey 1985; Sethi 1980), but quantity-insensitive for Tamil speakers (Vijaykrishnan 1978). This poses a challenge for researchers, since it is difficult to

anticipate stress patterns in materials. Nevertheless, most varieties involve prominent pitch movements on stressed syllables, wherever the latter occur (Pickering and Wiltshire 2000; Wiltshire and Moon 2003; Puri 2013).

Féry et al. (2016) analyse the narrow focus patterns produced by Hindi background speakers in terms of phrasal tones. However, studies by Maxwell (2014, Maxwell and Fletcher 2014) involving Bengali and Kannada speakers showed that for rising contours (i) peak delay relative to syllable landmarks did not vary with number of post-accentual syllables, and (ii) peak delay and L-to-H rise time were correlated with syllable duration. Since these results indicate peak alignment with stressed syllables and not the right edge of the word, they disfavour a phrase tone (i.e. edge tone) analysis. Studies nevertheless show that IE has at least two levels of phrasing (ip and IP, as in MEVs), with differential final lengthening (Puri 2013; Maxwell 2014).

Based on detailed timing data, Maxwell (2014) and Maxwell and Fletcher (2014) characterize the alignment characteristics of rising pitch accents in IE. For all speakers, L is consistently anchored to the onset consonant of the stressed syllable. For Kannada speakers, H aligns to the end of the accented vowel if it is long, and to the onset of the post-accented syllable if the accented vowel is short. For Bengali speakers, H aligns to the postaccentual vowel in nuclear focal accents, and variably to the accentual or postaccentual vowel in prenuclear accents. These results suggest that Kannada speakers have a single rising accent, while Bengali speakers use both L+H* and L*+H for prenuclear accents. ToBI analysis of read and spontaneous speech showed evidence for other pitch accent categories, including H* and L*, as well as downstepped variants of pitch accents and phrase accents. A study by Wiltshire and Harnsberger (2006) similarly found broad differences between Gujarati speakers, who use mostly rising pitch accents (L+H* or L*+H), and Telegu speakers, who additionally produce falling accents (H+L*, H*+L, H*).

In general, focused words are realized with some combination of greater duration, amplitude, and pitch excursion on the accented syllable (Moon, 2002; Wiltshire and Harnsberger 2006; Maxwell 2010; Maxwell 2014). Studies also report compression in the postfocal region (Féry et al. 2016) or alternation between deaccenting and post-focal compression without deaccenting (Maxwell 2014). Focus may also be marked by the insertion of a phrase boundary before or after the focused constituent (Maxwell 2010; Féry et al. 2017; Maxwell 2014).

Many regional and L1 varieties of IE remain understudied, thus more research is needed to corroborate the impression that IE varieties involve a similar phonological organization for intonation. Additionally, detailed phonetic evidence is needed to clarify many existing claims.

19.4.5 South Pacific Englishes (Niue, Fiji, Norfolk Island)

The islands of the South Pacific are home to a range of contact varieties of English. Because of patterns of economic migration, many of these varieties have greater numbers of speakers outside the islands (e.g. Niuean English, which has more speakers in New Zealand than on Niue). In addition, the movement of indentured labourers from India to Fiji in the late 19th and early 20th centuries has resulted in an Indo-Fijian English variety alongside Fijian English.

Starting with Fiji, since English is a second or third language for nearly all Fiji Islanders, there is L1 transfer in suprasegmental as well as segmental features (Tent and Mugler 2008). Stress patterns tend to be quite different from Standard English (i.e. in this case SSBE), such as [ˈkɒnˌsidɐret] for *considerate* or [ˌɛˈmikabɐl] for *amicable*. There is also a tendency for the nucleus to occur on the verb, even in unmarked sentences, such as *I am STAYing in Samabula*. The most marked intonational feature, however, is an overall higher pitch level

than in MEVs, especially a pattern on yes/no questions that starts high and ends with a rapid rise and sudden fall in pitch. This pattern (see Figure 19.4) is much closer to Fijian than to Standard English, and sometimes results in misunderstandings, where Standard English listeners have the impression that the speaker expects a positive response.

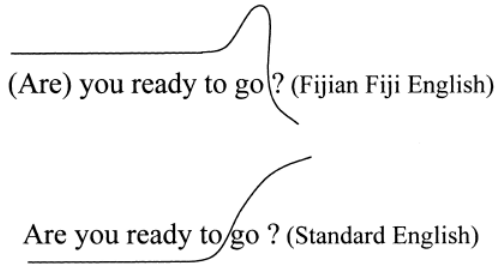

Figure 19.4. Intonation patterns for yes/no questions in Fijian and Standard English (from Tent and Mugler (2008: 249))

Durational differences between stressed and unstressed syllables are weaker in Niuean English than in MEVs (Starks et al. 2007). This variety also exhibits the high rising terminal or uptalk intonation pattern found in New Zealand English.

The status of Norfolk Island English ('Norfuk') as either a variety of English or as a creole is uncertain (Ingram and Mühlhäusler 2008). Its intonation is characterized as having a wide pitch range with much (often exaggerated) variation in pitch and tempo.

19.4.6 East African Englishes (Kenya/Uganda)

The intonation of East African Englishes is understudied. Otundo's (2017a,b) important groundwork reveals considerable differences in the English spoken by different L1 groups in Kenya, which she refers to as ethnically marked varieties. For declarative questions, Nandi speakers produce a rising pitch accent followed by a boundary rise (L*+H H%) whereas Bukusu speakers produce a high pitch accent followed by a fall (H* L%). Statements predominantly have either a rise-fall (L*+H L%) or fall (H* L%) in both groups, depending on the presence or absence of prenuclear material, respectively. WH-questions have a nuclear pitch accent on the WH-word, unlike MEV varieties, but they may have a terminal rise, in which case the nuclear accent is on a later word in the phrase.

Nassenstein (2016), in an extensive overview of Ugandan English, gives no detail on intonation, but does identify an interesting use of an intonational rise accompanied by additional vowel lengthening to mark assertive focus, as in (2).

(2) And he went ↑far up to there.
'And he went far (further than I had imagined).' (2016: 400)

Both Kenyan English and Ugandan English show less durational reduction of unstressed syllables compared to most MEVs (Schmied 2006).

19.4.7 Caribbean English

This section must begin with a caveat – descriptions of Caribbean English do not always clarify whether they are describing Caribbean-accented varieties of English or English-based Caribbean creoles. A notable exception is Devonish and Harry (2008), who explicitly distinguish Jamaican Creole and Jamaican English, identifying the latter as an L2 acquired through formal education by speakers of Jamaican Creole. The two languages coexist,

therefore, in a diglossic situation. The prosodic feature that these authors highlight is the prominence patterns on disyllabic words. For example, Jamaican Creole distinguishes between the kinship and priest meanings of *faada* ('father') by having a high tone on the first syllable[4] for the first meaning but on the second syllable for the second meaning, also treated as a lexical stress contrast /ˈfaada/ vs. /faaˈda/ (Cassidy and LePage 1967). In Jamaican English, these tonal patterns are maintained on /faaðo/, but lexical stress is on the first syllable in both cases.

Wells (1982: 572-574) highlights several other features of the prosody of West Indian English. These include a reduced durational distinction between stressed and unstressed syllables as compared to MEVs (see also Childs and Wolfram (2008)). At the same time, however, pitch ranges tend to be wider, which compensates somewhat for the reduction in the stress contrast by increasing the difference between accented and unaccented syllables. Wells also observes a tendency to shift stress rightwards, particularly in emphasis, giving /kɪˈʧɪn/ for *kitchen* (noted also for Eastern Caribbean creoles by Aceto (2008)), which he suggests might in fact be a rise-fall nucleus associated with the initial syllable, in autosegmental terms: L* H- L%. In this case rather than stress being shifted, the second syllable may simply have pitch prominence from the H- phrase accent.

A further intonational characteristic highlighted for Bahamian English (Childs and Wolfram 2008; Wells 1982) as well as Trinidadian and Tobagonian Creole (Youssef and James 2008) is the use of high rising contours with affirmative sentences. These rises appear to be inviting a response from the listener.

19.4.8 Black South African English

Given that English in South Africa is less common (at 9.6%) as a first language (L1) than IsiZulu (22.7%) or IsiXhosa (16.0%) and is the L1 for just 2.9% of Black South Africans (Statistics South Africa, 2011), it is unsurprising that Black South African English (BlSAfE) is heavily influenced by its speakers' L1s. Swerts and Zerbian (2010) compared intonation in the English of intermediate and advanced speakers who had Zulu as their L1. Both groups used rising and falling intonation patterns common to both Zulu and English to mark the difference between non-final and final phrases respectively, but only the advanced speakers patterned with the native L1 speakers of English in using intonation to mark focus (Zerbian 2015).

Coetzee and Wissing (2007) report that compared to White South African English and Afrikaans English, BlSAfE (in this case Tswana English) has less of a distinction in duration between stressed and unstressed syllables, and furthermore, does not show phrase-final lengthening. This supports similar general observations for BlSAfE by van Rooy (2004), who also notes that – again largely in line with the speakers' L1 – stress assignment is on the penultimate syllable (e.g. *sevénty*) unless the last syllable is super-heavy. This author also observes 'more frequent occurrence of pragmatic emphasis, leading to a different intonation structure of spoken BlSAfE' (p178) and notes that intonation phrases tend to be shorter than in White South African English (cf. Gennrich-de Lisle 1985).

19.4.9 Maltese English

Maltese English (MaltE), alongside Maltese and Maltese Sign Language, is an official language of Malta, with the great majority of its population being bilingual to different degrees.

[4] In other analyses of Caribbean creoles (e.g., Sutcliffe 2003) it is suggested that founder speakers of the creoles might have reinterpreted English word stress along lines of the tonal distinctions of West African languages, and realized stressed syllables with a high tone.

Maltese English does not reduce unstressed syllables to the same extent as MEVs and does not have syllabic sonorants in unstressed syllables (Grech and Vella 2019; cf. also chapter 16 this volume). As in a number of other contact varieties, Maltese English also differs from MEVS in stress assignment in compounds, such as *fire 'engine* and *wedding 'present,* with stress on the final rather than initial element, except in cases where the final element is monosyllabic, e.g. 'fireman (Vella 1995). However, Like Maltese, MaltE has regular pitch accents associated to lexically stressed syllables and tones associated with the right edge of intonation phrases. In WH-questions and some other modalities (e.g. vocatives and imperatives), tones can also be associated to the left edge of consituents (Vella 2003; 2012; Grice, Vella and Bruggeman 2019).

The right hand phrasal edge tone, a phrase accent in the sense of Grice, Ladd and Arvaniti (2000), is particularly important in the tonal phonology of Maltese English, and is often associated with a lexically stressed syllable (Vella 1995, 2003), leading to a percept of strong postfocal prominence (Galea Cavallazzi 2004)

19.5 Uptalk

A frequently discussed feature of English intonation is uptalk, sometimes referred to as 'high rising terminal' intonation (but see discussion in Warren (2016) and references therein). This use of rising intonation at the end of a declarative utterance should not be confused with the Urban Northern British rises, from which it is phonetically and functionally distinct. The term Antipodean Rise reflects its possible provenance in Australia and/or New Zealand, where it is becoming an unmarked feature. It is however found in many English varieties, including those spoken in the United States, Canada, South Africa, and the United Kingdom (see, e.g. Armstrong and Vanrell 2016; Arvaniti and Atkins 2016; Moritz 2016; Prechtel and Clopper 2016; Warren 2016; Wilhelm 2016).

Because of phonetic similarity to rises on yes/no and echo questions, uptalk is frequently interpreted by non-uptalkers as signalling that the speaker is questioning the content of their own utterances and is therefore uncertain or insecure. However, the distribution of uptalk and its interpretation by uptalkers indicate that it is used as an interactional device, to ensure the listener's engagement in the conversation. (For further discussion of the meanings of uptalk see Tyler and Burdin 2016; Warren 2016.)

In AM terms, uptalk has been labelled as L* H-H%, L* L-H% and H* L-H% for Canadian English (Di Gioacchino and Crook Jessop 2011; Shokeir 2007, 2008), L* L-H%, L* H-H% and H* H-H% for American English (Hirschberg and Ward 1995; McLemore 1991; Ritchart and Arvaniti 2013), L* H-H%, H* H-H% and the longer sequence H* L* H-H% for AusE and NZE (Fletcher 2005; Fletcher et al. 2005; Fletcher and Harrington 2001; McGregor and Palethorpe 2008), H* L-H% or H*+L H-H% for British English (Bradford 1997). This mix of labels indicates that there is variation both within and between varieties in terms of the shape of the uptalk rise, many labels including a falling component before the rise, so that the rise is from either a low pitch accent or a low phrase accent.

Recently, researchers have identified phonetic distinctions between uptalk and question rises, including the relative lateness of the rise onset in uptalk (in NZE, Warren 2005 and WSAfE, Dorrington 2010) and a lower rise onset in uptalk (especially in AusE, Fletcher and Harrington 2001); see Figures 19.5 and 19.6 respectively. The differences between uptalk and question rises can be perceived and interpreted appropriately by native speakers of the varieties (Fletcher and Loakes 2010; Warren 2005, 2014).

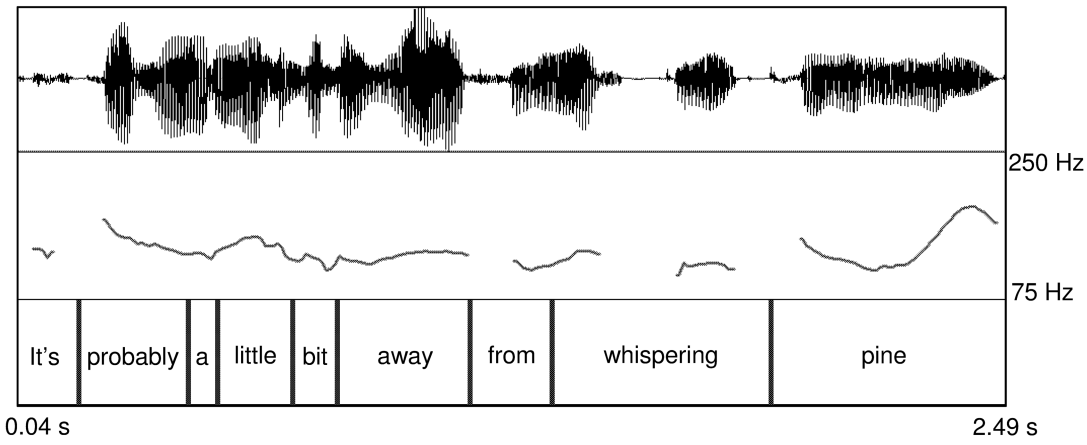

Figure 19.5 Fall-rise uptalk contour, Australian English (Warren 2016: Fig 4.1)

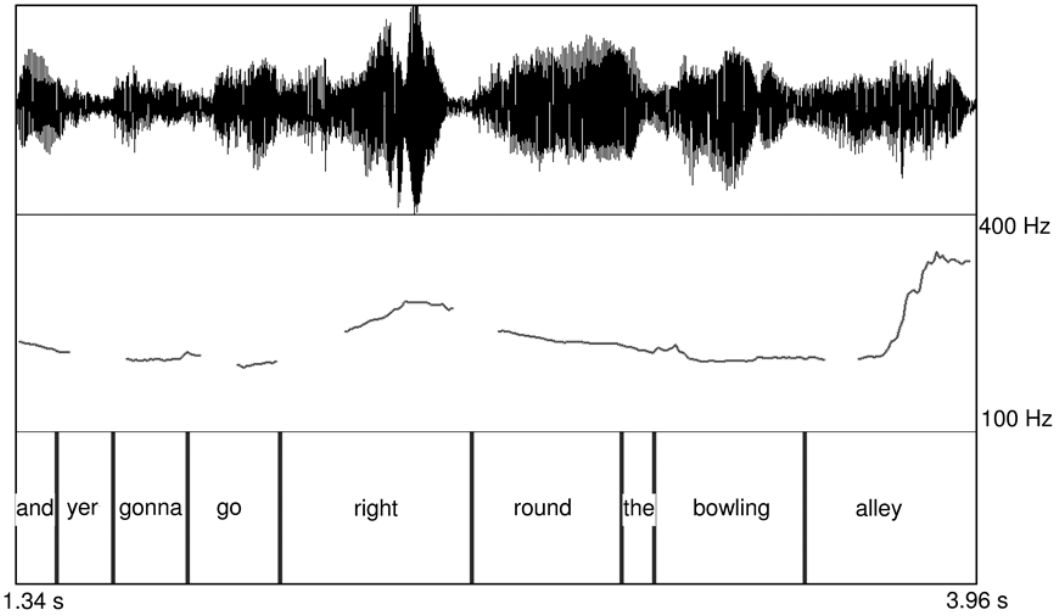

Figure 19.6. Late rise uptalk contour, New Zealand English (Warren 2016: Fig 4.2)

19.6 Conclusion

Our survey reveals a diverse set of intonational systems across varieties of English. For some - especially mainstream - varieties, diversity is limited to relatively minor features in inventories or in correspondences between contours and meanings. However, differences in overall phonological organization were also observed. There are, alongside stress-accent varieties, those like HKE, NigE, and Ghanaian English, which involve level tones assigned by both lexical specification and spreading rules. In terms of phrasing, MEVs and most other varieties involve a lowest level of phrasing that is relatively large, while SgE patterns with languages like French and Korean in having smaller units of phrasing (i.e., accentual phrases) that generally span only one or two content words. At the other end of the spectrum, HKE, NigE, and Ghanaian English include only one large unit of phrasing (i.e., intonation phrases) that contribute to the tonal specification. In light of issues raised by MEVs concerning the linking of tones to either pitch accents or edge tones (see §19.3.1), an important issue for future research is whether the need for specific phrasing levels can be supported by evidence that is independent of the edge tones they are purported to carry (e.g., progressive final lengthening).

All the English varieties we cover include word stress in some form or another. This can involve a manifestation of greater acoustic prominence, or else merely designates syllables as prosodically privileged. In the latter case, privileged syllables are singled out to receive specific tonal values without concomitant acoustic or articulatory enhancement, and in most cases, these tones are lexically specified. The high functional load of pitch accents in mainstream varieties has most probably led to the need for the location of these pitch accents, lexical stress, to be reinterpreted as tone.

Across the different varieties, there is considerable variation in the assignment of word stress, differing from MEVs for certain words (Fijian, where secondary stress and primary stress are sometimes exchanged, AAE, where there is a preference for initial stress) or word types (the compound rule is different in Maltese and Hispanic English). In Indian English, stress is rule-based, with quantity sensitivity in some L1 groups (e.g. Hindi or Punjabi), but overall the rules are highly variable. In BlSAfE stress also appears to be rule-based, with penultimate stress in most cases. Caution is required, however: Sometimes it might appear that word stress is shifted when in fact the pitch accent is low and is followed by a H phrase accent (Caribbean E, SgE), or when constituents in a compound are treated as separate word domains (NigE, GhanE). A similar issue applies to sentence-level prominence (and by extension, word citation forms), since in varieties like NigE or HKE, a lack of postnuclear deaccenting, combined with rightward spreading of H tones, can give the impression that prominence is 'shifted' to a later word or syllable than in MEVs. It is therefore essential that the interplay of stress, accent, and tonal assignment be worked out separately for each variety and independently of what might give the impression of prominence in another variety.

This diversity highlights the important role played by the choice of framework and analytical tools when approaching any variety, whether unstudied or well-studied. It would, for example, be clearly problematic to apply the notion of a nuclear contour to HKE, NigE or GhanE, since, in those varieties, the tune at the end of a phrase is fully determined by lexical word stress and the choice of boundary tone. Apart from the latter, in other words, there is no optionality that could result in meaningful contrasts (see Gussenhoven [2017] for further discussion). Additionally, we need look no further than MEVs to recognize that aspects of the prenuclear contour can be informative in terms of both structure and meaning. Thus, applying autosegmental-metrical or British school categories developed for a well-studied variety to a variety whose phonological organization and intonational inventory has not yet been established is highly problematic, since it strongly biases the researcher towards an analysis in terms of pitch accents (or even nuclear pitch accents), thereby missing other typological possibilities.

Moreover, even if the variety in question does pattern with MEVs in having regular pitch accents, the use of a pre-existing inventory runs the risk of (i) treating non-contrastive differences as contrastive, and (ii) lumping existing contrasts into a single category. This issue is underscored by the numerous studies on Urban Northern British and American English varieties (Arvaniti and Garding 2007; Clopper and Smiljanić 2011) as well as IE (Maxwell 2014; Maxwell and Fletcher 2014). In these cases, detailed phonetic studies revealed inventory differences across varieties of these languages that would have been missed had the authors been restricted to any of the inventories available for MEVs.

Besides the differences outlined above, our survey also revealed certain regularities. For example, the majority of non-MEV varieties either lack postfocal deaccenting, or use it with less regularity. These include HKE, SgE, NigE, GhanE, BlSAfE, IE, MaltE, and AAE. This broad tendency suggests that post-focal deaccenting as a feature tends to resist transfer in contact situations, either because it is not compatible with certain types of phonological organization, or because it is not sufficiently salient to L2 speakers during nativization.

It is also interesting to note the striking similarity in the intonation systems of HKE and West African varieties, which resulted from contact with unrelated tone languages. As Gussenhoven (2017: 22) notes, this can be 'explained by the way the pitch patterns of BrE words were translated into tonal representations'. While some accounts suggest that a similar generalization applies to SgE, that variety has received substantial influence from intonation languages including Malay, Indic and Dravidian languages, and even IE (Gupta 1998), which could explain why it patterns less closely with the above 'tonal' varieties.

In this chapter we have only been able to report on a selection of contact varieties of English. The typological diversity we have observed will no doubt be enriched, once we take up the challenge of analysing newly emerging (extended circle) varieties, such as those spoken in China and Korea.