

Chapter 5. Prosodic fieldwork

by Nikolaus P. Himmelmann and D. Robert Ladd

This chapter is a little different from the preceding and following chapters, because it devotes a lot of space to basic issues that have nothing directly to do with fieldwork. The reason for this is simple. There's a lot of basic agreement about how segmental phonology and morphosyntax work, and most linguists know what kinds of things they will be looking for when they begin work on a new language. So in chapter 5 and 7 the focus is not on what to look for, but how to look. But with prosodic features – the kinds of things that often don't show up in a segmental transcription – many linguists feel that they are on shaky ground. They are insecure about hearing prosodic distinctions and unclear about the way these distinctions might be used in different languages. This means that an important goal of this chapter is to make you aware of what you might be looking for – not just how to look.

There are two fundamental ways that prosodic features differ from more familiar segmental features. One is that they are relevant at different levels of structure: there are both word-level or lexical prosodic features and sentence-level or “post-lexical” ones. Probably the best known typological difference based on this distinction is the one between “tone languages” like Chinese, where pitch serves to distinguish otherwise identical lexical items, and non-tonal languages like English, where pitch only serves to signal sentence-level differences of “intonation”. However, the use of prosodic features at different levels applies more widely as well: in English we can use stress at the lexical level to distinguish one word from another (e.g. *PERmit* [noun] and *perMIT* [verb]), but also at the post-lexical level to distinguish one sentence meaning from another (e.g. *I only put salt in the STEW* and *I only put SALT in the stew.*)

The other important property that sets prosodic features apart from familiar segmental features is that their sentence-level functions – like intonation and sentence-stress – are often broadly similar even in completely unrelated languages. For example, it is very common cross-linguistically to signal questions by the use of sustained high or rising pitch at the end of an utterance, even in languages that also have lexical tone. Because it works at different levels and because it has both universal and language-specific aspects, prosody is likely to seem mysterious and difficult. Speakers of a language that uses a given feature in one way are likely to find using it in a different way strange and exotic and (more practically) hard to hear: this is a common reaction of speakers of non-tonal languages when they encounter a tone language. Furthermore, sentence-level distinctions are probably inherently more difficult to think about than lexical distinctions: the difference between a pin and a bin is instantly obvious and easy to demonstrate, whereas the difference between the two versions of the sentence about the salt and the stew in the preceding paragraph takes careful explaining.

Nevertheless, prosody is an essential ingredient of every spoken language, and a description of prosody is an essential ingredient of every complete language description. In the following sections we will sketch some of the key phonetic, functional, and typological aspects of prosodic features, then go on to outline various techniques for achieving a satisfactory analysis of prosodic features in the field.

5.1 The phonetic fundamentals

We begin by briefly introducing four phonetic parameters which are relevant to prosody: pitch, duration, voice quality, and stress.

5.1.1 Pitch

Pitch is the property that distinguishes one musical note from another. In speech, pitch corresponds roughly to the fundamental frequency (F0) of the acoustic signal, which in turn corresponds roughly to the rate of vibration of the vocal cords. It is physically impossible to have voice without pitch – if the vocal cords are vibrating, they are necessarily vibrating at some frequency. In English and many other European languages we talk about pitch being “higher” or “lower” as the frequency of vibration gets faster or slower, but other sensory metaphors are used in other languages and cultures (“brighter/darker”, “sharper/duller”, etc.). Perhaps because pitch is a necessary property of voice, all languages – so far as we know – exploit pitch for communicative purposes.

The most striking thing about pitch is that it varies conspicuously from one speaker to another – men generally have lower voices than women. This means that the phonetic definition of pitch for linguistic purposes cannot be based on any absolute level of fundamental frequency but must be considered relative to the speaker’s voice range. Normalization for speaker differences must also deal with the fact that speakers can “raise their voice” without affecting the linguistic identity of pitch features. The details of how this normalization should be done are not fully clear but the basic principle is not in doubt. Moreover, this seldom causes serious practical difficulties in fieldwork, because we can usually hear whether a given pitch is relatively high or low in the speaker’s voice.

However, even if we find it relatively easy to abstract away from differences of overall pitch level, there are still major difficulties in the phonetic description of pitch. This is reflected in the lack of any agreement on an IPA system for transcription of pitch distinctions. One of the key issues for transcription is the relevance – or lack of relevance – of the syllable. In many tone languages, terms like “rise” and “fall” must be defined *relative to the syllable*: a sequence of a high-tone syllable and a low-tone syllable can be lexically completely different from a sequence of a falling-tone syllable and a low-tone syllable, even though both sequences involve an overall “fall” in pitch over the two syllables. In such a tone language, the overall “fall” is not relevant for phonetic description. In a language like English, on the other hand, a phonetic fall on a monosyllabic utterance (e.g. *John*) and a phonetic high-to-low sequence on a disyllabic one (e.g. *Johnny*) may be completely equivalent in the intonational system, which suggests that the “fall” must be regarded as a phonetic event regardless of the number of syllables it spans. This idea is strengthened by recent work showing that in languages like English and German functionally equivalent pitch movements can be “aligned” in different ways relative to syllables in different languages and language varieties.

In studying an unfamiliar language, in short, the fieldworker needs to be alert to the fact that descriptive assumptions can be hidden even in an apparently neutral label like “pitch fall”. For fieldwork, the most important thing to know about pitch is that a useful phonetic description of pitch depends on the way pitch is used in the language. More practically, the fieldworker whose native language works like English must be prepared to detect the syllable-by-syllable phonetic chunking of the pitch contour that is likely to be found in a language with lexical tone.

5.1.2. Segment duration

To the extent that we can divide an utterance into phonetic segments with clearly defined boundaries, we can measure the duration of the segments. In many languages duration is systematically manipulated for prosodic effect (e.g. distinctions between long and short vowels), but in all languages segment duration is affected by a whole range of other factors as well. These include some nearly universal allophonic effects (e.g. vowels tend to be longer before voiced consonants than before voiceless consonants; low vowels tend to be longer than high vowels; fricatives tend to be longer than stops) and effects of speaking rate (faster rate means shorter segments, but vowels are generally more compressible or expandable than consonants). Segment duration is also affected by other prosodic factors: specifically, stressed vowels tend to be longer than unstressed vowels; segments in phrase-final syllables tend to be longer than in other syllables; and word-initial and phrase-initial consonants tend to be longer than consonants in other positions.

For fieldwork, these differences mean that any suspected duration distinctions must always be checked in similar sentence contexts. In particular, if you ask someone to repeat two items that appear to be a duration-based minimal pair (like *Stadt* ‘city’ and *Staat* ‘state’ in German), it is important to hear the two members of the pair in both orders. That way you will not be misled by any lengthening (or occasionally, shortening) of whichever item is pronounced second.

5.1.3. Voice quality

The phonetic description of voice quality is less well advanced than that of other prosodic features. Many differences of voice quality – described by such impressionistic terms as “harsh”, “breathy”, “creaky” and so on – are based on different configurations of the glottis. As such they are difficult to observe directly, either in ourselves or in others, except by the use of special equipment. The standard work on the impressionistic description (and transcription) of voice quality is Laver 1980, which remains a useful reference for fieldwork. Much recent research has focused on understanding the acoustic correlates of voice quality differences and/or the glottal configurations that give rise to them. This work is not likely to be of much direct relevance to descriptive fieldwork, but good fieldwork can provide the basis for directing instrumental phonetic studies into fruitful areas of research.

5.1.4. Stress

Roughly speaking, stress is the property that makes one syllable in a word more prominent than its neighbours – for example, signaling the difference between the noun *PERmit* and the verb *perMIT*. Perhaps surprisingly, it is extremely difficult to provide a phonetic definition for this “greater prominence”. Impressionistically (for native speakers of most European languages), the phonetic basis of stress is “loudness” – the stressed syllable seems louder than neighbouring unstressed syllables – but perceived loudness is psychophysically very complicated, not just in speech but in all auditory stimuli. The most important correlate of loudness is intensity (sound energy), but duration and fundamental frequency have also been shown to play a role – for the same peak intensity, a longer or higher-pitched sound will sound louder than a shorter or lower-pitched one.

A possibly more useful phonetic definition of stress is “force of articulation”, which shows up less in effects on the overall energy in a segment or syllable and more in the distribution of energy in the spectrum of the sound. Specifically, it has recently been suggested that stressed vowels in Dutch have more energy at higher frequencies than unstressed vowels (they have “shallower spectral tilt” [Sluijter and van Heuven 1996]). There may also be effects of “force of articulation” on the relative duration of consonant and vowel portions of a syllable, although the details are not at all clear. Additionally, stressed syllables usually contain full (peripheral) vowels while unaccented syllables may contain reduced (centralized) vowels such as schwa; alternatively, a language may have only or mainly peripheral vowels, but stressed syllables may allow for larger vowel inventories than unstressed syllables. For example, Catalan distinguishes seven vowels /i e □ a □ o u/ in stressed syllables but only three /i □ u/ in unstressed ones.

Whatever the phonetic basis of stress turns out to be, and however much it turns out to differ from one language to another, it is clear that part of the problem defining the phonetic basis of “stress” is the existence of conceptual and theoretical problems with the classification and description of accentual systems generally. We return to this issue in the next section, and in section 6.3.3.

5.2. Typical functions of prosodic features

5.2.1. Lexical and morphological functions

The lexical functions of prosody are, on the whole, like the function of most segmental phonological distinctions: to distinguish between one lexical item and another. Just as English *pin* and *bin* differ minimally phonologically but are two unrelated lexical items, so pairs like Chinese *niàn* ‘study’ and *nián* ‘year’ or Dutch *man* ‘man’ and *maan* ‘moon’ or Greek [□□□□□] ‘space’ and [□□□□□] ‘dance’ involve unrelated lexical items that are minimally different phonologically. Similarly, just as segmental distinctions can be used to signal different morphological categories (for example, English *foot/feet* for singular/plural or *drink/drank* for present/past), so prosodic features can be used in the same way, as in the difference between Efik [□□□□□] ‘I buy’ and [□□□□□] ‘I would buy’ or Dinka [dèk] ‘drink (infinitive)’ and [dè:k] ‘s/he drinks’ or Italian [□□□□□] ‘I speak’ and [□□□□□] ‘s/he spoke’.

The examples just given illustrate the three most commonly encountered types of lexical prosodic distinctions: **tone** (as in the Chinese and Efik examples), **quantity** (as in the Dutch and Dinka examples) and **accent** (as in the Greek and Italian examples). It is common to treat the three of these together as “suprasegmental” features, and to identify them with the phonetic parameters of **pitch**, **duration** and **stress**. A classic statement of this view, still useful for the data it contains, is Lehiste’s book *Suprasegmentals* (1970). However, this view is misleading in two distinct ways. First, the linguistic categories of tone, quantity and accent are often cued in multiple phonetic ways. Tone is primarily a matter of pitch, but may also involve accompanying differences of segment duration and voice quality: for example, in Standard (Mandarin) Chinese syllables with “tone 3” are not only low in pitch but tend to be longer in duration and to have creaky or glottalised voice as well. Quantity distinctions are based on segment duration, but often involve differences of vowel quality or (in the case of consonants) manner of articulation as well: for example, German long vowels tend to be higher and less central than their short counterparts, in addition to being longer in duration. As for accent, there are so many different phonetic manifestations of things that have been called “stress” or “accent” that there is very little agreement on what these terms refer to. In short, it is at best a gross oversimplification to think of tone, quantity and accent as the linguistic functions of the phonetic features pitch, duration and stress.

The second reason for not treating tone, quantity and accent together is that they are functionally quite different. Where they exist, distinctions of tone and quantity are often functionally similar to segmental distinctions. Tone – especially in East Asia, West Africa, and parts of the Americas – generally has a high functional load, and it is not at all uncommon to find extensive minimal sets distinguished only by tone, for example Yoruba *igba* ‘two hundred’, *ìgbá* ‘calabash’, *ìgbà* ‘[type of tree]’, *ìgbà* ‘time’. Quantity systems are similar: in many languages with distinctive vowel or consonant quantity, all or almost all the vowels or consonants can appear both long and short in pairs of unrelated words, for example German *bitten* /bitn/ ‘request’ vs. *bieten* /bi:tn/ ‘offer’; *beten* /be:tn/ ‘pray’ vs. *Betten* /betn/ ‘beds’, etc.

By contrast, accentual differences are often rather marginal in the lexicon of a language as a whole, yielding few minimal pairs and/or involving some sort of morphological relatedness. For example, in English the lexical accent in a word is certainly a distinctive part of its phonological make-up, and a misplaced stress (e.g. in foreign pronunciation) can make word identification very difficult. Yet there are very few minimal pairs in English based on lexical accent, except for derivationally related noun-verb pairs like *OBject-obJECT* and *PERmit-perMIT*. This difference is due to the fact that accent involves a syntagmatic relation (the relative prominence of two syllables) whereas tone and quantity, like most segmental features, are a matter of paradigmatic contrasts between members of a set of possible phonological choices. It is clearly meaningful to say of a monosyllabic utterance that it has a long vowel or a high tone, because these terms can be defined without reference to other syllables. It is often less clear what it means to say that a monosyllabic utterance is “stressed” or “accented”.

Finally, we should mention lexical distinctions of voice quality, which are often not considered under the heading of “prosody” at all. In some languages there are

phonemic distinctions of voice quality which are associated with specific consonantal contrasts: for example, in Gujarati voice quality distinctions are historically related to the distinction between “voiced” and “voiced aspirated” stops in other languages of Northern India, and are found only in the presence of certain specific stops. Similarly, in many East Asian tone languages there are characteristic differences of voice quality that accompany pitch differences in distinguishing between one tone phoneme and another, and which are therefore generally described as part of the tonal system. (This is the case with the glottalisation that often accompanies Mandarin “Tone 3”, as we just saw above.) However, voice quality distinctions can be independent of both segmental and tonal distinctions: for example, in Dinka, the two distinctive voice qualities (often described as “creaky” and “breathy”) can cooccur with any of the tone phonemes, any of the distinctive quantity categories, and most of the vowel and consonant phonemes.

5.2.2 Sentence-level functions

At sentence level, prosodic features typically play a role in marking three general functions: sentence modality and speaker attitude, phrasing and discourse segmentation, and information structure and focus. However, there is nothing intrinsically “prosodic” about any of these functions: all of them may also be marked in a non-prosodic way in addition to, or instead of, a prosodic marking. Thus, for example, while sentence modality and focus are often marked by intonational means in many European languages, many other languages employ particles or affixes in the same functions (e.g. focus particles in Cushitic languages, question marking clitics in western Austronesian languages).

An important problem in studying the prosodic signaling of these functions is that many pitch-related phenomena are quasi-universal, which reflects their link to prelinguistic ways of communicating that we share with other species. As noted in section 6.1.1, women have higher-pitched voices than men, and individuals can “raise” and “lower” their voices for various expressive purposes. These “paralinguistic” functions of pitch and voice quality are broadly similar the world around, though there are big differences between cultures in the way the paralinguistic functions are evaluated. For example, a voice raised in anger sounds much the same in any language, but raising the voice in that way may be dramatically less acceptable in one culture than in another. Similarly, in some cultures it is highly valued for males to have very low voices and/or for females to have very high voices, and speakers tend to exaggerate the biologically based differences, whereas in other cultures little importance is attached to such differences.

5.2.2.1 Sentence modality and speaker attitude

The prosodic expression of modality and attitude is most closely identified with speech melody and voice quality. Together, these are the characteristics we are most likely to think of as the “intonation” of an utterance. Typical examples include the use of overall falling pitch in statements, overall rising pitch in yes-no questions, or the use of overall high pitch in polite utterances.

These examples are also typical examples of the difficulty of distinguishing linguistic and paralinguistic functions of pitch. For example, there have been disagreements about

whether overall rising pitch in “question intonation” is part of a language-specific intonational phonology or merely based on the universal use of high pitch to signal tentativeness or incompleteness. Our view is that it is necessary and appropriate to talk of “intonational phonology” for at least some sentence-level uses of pitch (see further section 6.3.1 below). It is important to remember that languages may diverge considerably from the quasi-universal tendencies mentioned above: there are languages such as Hungarian or some dialects of Italian, where question intonation includes the kind of final fall which is typical of statements in other western European languages. Nevertheless, we acknowledge that there is genuine empirical uncertainty about how to distinguish phonologized uses of pitch from universal patterns of human paralinguistic communication.

5.2.2.2 Phrasing and discourse segmentation

In all languages, so far as we know, longer stretches of speech are divided up into prosodically defined chunks often called intonation units (IUs) or intonation(al) phrases (IPs). To some extent this division is determined by the need for speakers to breathe in order to continue speaking, and in the literature the term breath group may also be found for what we are here calling IU. However, it is important not to think of IUs purely as units of speech production, because they almost certainly have a role in higher-level linguistic processing as well, both for the speaker and the hearer. That is, intonation units are also basic units of information (e.g. Halliday 1967, Chafe 1994, Croft 1995) or of syntax (e.g. Selkirk 1984, Steedman 2000). Closely related to the issue of segmentation into IUs are the prosodic cues that help control the smooth flow of conversation (e.g. signals of the end of one speaker’s turn) and the cues that signal hierarchical topic structure in longer monologues such as narratives (e.g. “paragraph” cues). An eventual theory of prosodic phrasing will cover all these phenomena.

The phonetic manifestations of phrasing and discourse chunking are extremely varied. The clearest phonetic marker of a boundary between two prosodic chunks is a silent pause, but boundaries can be unambiguously signaled without any silent pauses, and not all silent pauses occur at a boundary. Other cues to the presence of a boundary include various changes in voice quality and/or intensity (for example, change to creaky voice at the end of a unit), substantial pitch change over the last few syllables preceding the boundary (such as an utterance-final fall), pitch discontinuities across a boundary (in particular, “resetting” the overall pitch to a higher level at the beginning of a new unit), and marked changes in segment duration (especially longer segments just preceding a major boundary). However, it is also important to note that there are extensive segmental cues to phrasing as well, especially different applications of segmental sandhi rules. For example, in French, “liaison” – the pronunciation of word-final consonants before a following vowel – is largely restricted to small phrases and does not occur across phrase boundaries: *allons-y* ‘let’s go’ (lit. ‘let’s go there’) is pronounced [alɔ̃zi] but *allons à la plage* ‘let’s go to the beach’ is normally pronounced [alɔ̃alaplɑ̃], signalling the presence of a boundary between *allons* and *à la plage*.

An important conceptual problem in discussing phrasing and discourse segmentation is that we need to recognize different levels of prosodic structure, and there is no agreement on how to do this. In corpora of ordinary spontaneous speech it will often be easy enough to distinguish a basic level of IU, perhaps 6-10 syllables long, set off

by relatively clear boundaries signaled by silent pauses and other cues. However, merely dividing texts into a single level of IUs tells us nothing either about the smaller units that distinguish one syntactic structure from another, nor about the larger units (often called “episodes” or “paragraphs”) that signal higher-level textual organization in monologues. This important topic is unfortunately beyond the scope of this chapter.

5.2.2.3 Information structure and focus

Related to the marking of boundaries and cohesion is the use of prosody to signal semantic and pragmatic features often collectively known as “information structure”. This includes notions like “contrast”, “focus” and “topic”, and refers to the way new entities and new information are introduced into a discourse and to the way in which entities and information already present in a discourse are signaled as such. One important means of conveying this kind of information is to put specific words or phrases in prosodically prominent or non-prominent positions. In some languages word order can be extensively manipulated in order to achieve this, whereas in other languages the same string of words can have different prosodic structures. Both strategies are exemplified in English constructions involving direct and indirect objects: we can say either *I gave the driver a dollar* or *I gave a dollar to the driver*, putting either the amount of money or the recipient in the prosodically prominent final position. Other things being equal, the first construction is used when the amount of money is more informative in the discourse context and the second when the point of the sentence is to convey something about the recipient. However, we can achieve similar effects by restructuring the prosody so that the major sentence-level prosodic prominence occurs on a non-final word: *I gave the DRIVER a dollar* (... not the waiter) or, somewhat less naturally, *I gave a DOLLAR to the driver* (...not a euro).

There is an extensive literature on these matters, especially in the European languages; the reader is referred to Lambrecht (1994) and Ladd (1996) for useful summaries. Fieldworkers should probably be wary of expecting to find close analogues of European phenomena in languages in other parts of the world.

5.3. Phonology of tone, intonation and accent

From the foregoing sections it will be clear that “prosodic” features – defined on the basis of phonetic properties that are not normally indicated in a segmental transcription – do not form a linguistically coherent set. Among other things, this means that there is no way of knowing ahead of time how the phonetic features loosely referred to as “prosodic” – pitch, duration, and so on – are going to be put to phonological use in any given language. Speakers of all languages produce and perceive differences in pitch, duration, voice quality, and probably relative prominence, but they may interpret these differences in radically different ways. There is no unique relation between a given phonetic feature and its phonological function.

As we suggested earlier, some “prosodic” distinctions turn out to work in ways that are no surprise to any linguist, while others – sometimes involving the same phonetic raw material – are still in need of extensive new theoretical understanding before we can be sure that our descriptions make sense. What seems fairly clear is that the “unsurprising” prosodic features (like lexical tone and quantity) involve linguistic

elements that are grouped into strings and contrast paradigmatically with other elements, like most segmental phonemes. The “problematical” prosodic features (like accent and phrasing) are somehow involved in signaling phonological structure, the grouping of linguistic elements into larger chunks. In this section of the chapter we provide a little more detail on two problematical topics: the tonal structure of intonation, and the nature of “accent”.

5.3.1 Tone and intonation

As we’ve already seen, pitch provides the main phonetic basis for prosodic distinctions both at the word level (“tone”) and at the sentence level (“intonation”). Tone languages are extremely varied, and it would be possible to devote this entire chapter just to describing the many varied phenomena of lexical and grammatical tone. However, since there are good descriptions of numerous prototypical tone languages from around the world and a substantial body of literature discussing various aspects of their analysis, it would be pointless to attempt a mere summary here. The textbook by Yip (2002) provides a comprehensive survey, and is a useful guide to various descriptive and theoretical problems. Anyone embarking on the study of a language known or suspected to have lexical and/or grammatical tone should be well acquainted with this literature before leaving for the field.

We focus here instead on intonation. We use the term here in a strict sense, to refer to sentence-level uses of pitch that convey distinctions related to sentence modality and speaker attitude, phrasing and discourse grouping, and information structure. The phonological structure of intonation is better understood now than it was a few decades ago, but there are undoubtedly plenty of intonational phenomena waiting to be discovered in undocumented languages, and plenty of things that we will understand better once we have a fuller idea of the range of possibilities. What we present here is a minimal framework for investigating intonation in a new language. Our discussion is based on the now widely accepted “autosegmental-metrical” theory of intonation (for a review see Ladd 1996).

The most important phonological distinction to be drawn is the one between intonational features at major prominent syllables and intonational features at boundaries: in current terminology, the distinction is between “pitch accents” and “boundary tones”. The existence of such a distinction has been recognized by some investigators since the 1940s, and is made explicit in current autosegmental-metrical transcription systems for numerous (mostly European) languages. The difference between the two can be readily appreciated in English when we apply the same intonational tune to sentences with markedly different numbers of syllables and/or markedly different accent patterns. For example, imagine two different possible astonished questions in response to the sentence *I hear Sue’s taking a course to become a driving instructor*. One might respond *Sue?! or one might respond A driving instructor?! In the first case, the pitch of the astonished question rises and then falls and then rises again, all on the vowel of the single syllable Sue. In the second case, the pitch is briefly fairly level at the beginning, then there is a steep rise in pitch on the lexically stressed syllable dri-, immediately followed by a fall, then a level low-pitched stretch until the very end of the utterance, at which point there is an abrupt rise. [Really need a sound file and a picture here!///] At a minimum, therefore, the contour consists*

of two separable parts: a rising-falling movement at the main stressed syllable and a rise at the very end. On the monosyllabic utterance *Sue* these two parts are compressed onto the single available syllable, which is both the main stressed syllable and the end of the utterance. But with a somewhat longer phrase the separateness of the two prosodic events becomes clear.

One important clue to the correctness of the distinction between pitch accents and boundary tones is the fact that in some lexical tone languages, where pitch primarily conveys lexical information, there are nevertheless intonational pitch effects at the ends of phrases or sentences. These effects typically involve modifications of the lexically-specified pitch contour on the pre-boundary syllable. Early descriptions of this effect were given by Chuang 1958 for Szechuan Mandarin and by Abramson 1962 for Thai. This coexistence of lexical and intonational pitch can be described easily if we recognize boundary tones: in these languages the pitch contour of an utterance is principally determined by the lexical tones of the words that happen to make it up, but at the edges of phrases it is possible to add an additional tonal specification - a boundary tone.

However, it should be emphasized that not all lexical tone languages use intonational boundary tones; for example, some West African tone languages appear not to have them, so that in these languages the pitch contour of an utterance is almost completely determined by the string of lexical tones. Conversely, there appear to be languages with intonational boundary tones that have neither pitch accents nor lexical tonal specifications. In these languages, all intonational effects are conveyed by pitch movements at the edges of phrases, and “nothing happens” phonologically in between. Obviously, there is phonetic pitch wherever there is voicing, but the linguistically significant pitch effects are restricted to phrase edges, and the pitch in between is determined by simple interpolation. Clear descriptions of such systems are given by Robert and Rialland (2001) for Wolof and Jun (1998) for Korean.

Current transcription systems for pitch accents and boundary tones, based largely on the ToBI system first designed for English in the early 1990s, analyze these pitch movements further: the astonished question contour just discussed would probably be transcribed as a L+H* pitch accent, an immediately following L- “phrase accent”, and a H% or L+H% boundary tone. The details are well beyond the scope of this chapter, but the reader who expects to deal with an unfamiliar intonation system in a language without lexical tone should consult the ToBI web site (URL <http://www.ling.ohio-state.edu/~tobi/>) and its extensive series of links to ToBI systems that have been designed for a number of other languages.

Before we leave the subject of intonation, we must note that in addition to pitch accents and boundary tones, intonation can make crucial use of what we might call “register effects”. Recall that the phonetic realization of pitch distinctions is somehow relative to the speaker’s pitch range: “high” does not refer to some absolute fundamental frequency level, but a level that is high for a given speaker in a given context. This even applies within a single utterance: as a result of the widespread phenomenon of “declination” – a gradual lowering of pitch across a phrase or utterance – the pitch of a “high” tone at the end of an utterance may be lower than that of a “low” tone at the beginning. That is, the phonological interpretation of pitch level is somehow relative to

a frame of reference that varies not only from speaker to speaker and from context to context but also from one part of an utterance to another. Such changes of the frame of reference during the course of an utterance can be exploited for communicative purposes in various ways, and these are what we are calling “register effects”. The clearest examples of such effects involve the interaction of lexical tone and overall pitch level to signal questions. In Chinese, for example, it is possible (though not very usual) to distinguish yes-no questions from statements in this way.

5.3.2 Lexical accent systems

The existence of tone languages is such a remarkable fact from the point of view of speakers of non-tonal languages that there are at least two typological schemes – devised by speakers of non-tonal languages – that attempt to accommodate lexical/grammatical tone in a larger theoretical understanding. One of these is based on the “domain” of pitch distinctions, while the other is based on a typology of “word prosody”. Looking at the domain of pitch, languages have been divided into “tone languages” (where the domain of pitch distinctions is the syllable), “melodic accent languages” (where the domain of pitch distinctions is the word), and “intonation languages” (where the domain of pitch distinctions is the phrase or utterance). This typology goes back at least to Pike (1945) and is found in work as recent as Cruttenden (1997). Looking instead at the lexical uses to which “prosodic” features are put, we can divide languages into “tone languages” (in which each syllable has different tonal possibilities), “melodic accent languages” (in which one syllable in a word or similar domain is marked by pitch in some way), and “dynamic accent languages” (in which one syllable in a word or similar domain is marked by stress in some way). This typology is suggested by Jun 2005. Both typologies have obvious problems (e.g. the existence of intonational distinctions in tone languages, the existence of languages like Swedish with both dynamic accent and lexically specified melodic accent), and neither commands wide acceptance.

In our view, the problems with these typologies result from trying to incorporate tone and accent in the same scheme. As we pointed out earlier, tone often functions like segmental distinctions: it involves a choice of categories from a paradigmatic set, and it is meaningful to talk about e.g. a high tone on one syllable without reference to the tone on any other syllable. Accentual distinctions, on the other hand, are syntagmatic distinctions: they involve contrast with immediately adjacent syllables in a string. Consequently, we believe that it is quite misleading to see, as in Pike’s typology, a continuum from tone to melodic accent to intonation, and equally misleading, but in a different way, to take “tone” and “stress” as different kinds of “word prosody” that a language may have. Rather, we think it will be useful to discuss the ways in which accentual systems can differ without necessarily trying to place them into a typological scheme that places them in the same dimension as intonation and tone (i.e. the typology of prosodic systems should involve three, at least partially independent dimensions: tone, accent, and intonation).

A general and possibly universally valid definition of lexical accent is the singling out of a specific syllable in a word or similar domain (such as the “foot”) for some sort of prominence or other special prosodic treatment. Lexical accent, as conceived of this way, is an abstract structural notion, and says nothing about how exactly the “special

prosodic treatment” is manifested in the acoustic signal. In some languages, the special status of the accented syllable is based entirely on association with a specific pitch feature; in other languages, the accented syllable is distinguished from other syllables by phonetic “stress” – greater force of articulation leading to some combination of longer duration, greater intensity, more peripheral vowel quality, shallower spectral tilt, etc. (cf. section 6.1.4). This suggests a distinction between “melodic” and “dynamic” accent, a traditional distinction recently reestablished by Beckman (1986).

The distinction between melodic and dynamic accent is a phonetic one. Other typological dimensions on which accentual systems appear to differ involve structural properties. These include obligatoriness, culminativity, recursivity, transitivity, intonational anchoring, and lexical distinctiveness. We briefly outline these six properties here:

Obligatoriness: In some accentual systems, an accent must occur within each domain of the specified size: if the “prosodic word” is the domain of accent, then each prosodic word must have an accent. In other systems, the accent may or may not occur in a given domain. For example, in Japanese, words can be accented or unaccented, whereas in English any word of more than one syllable must have at least one syllable that stands out as more prominent when the word is pronounced in isolation.

Culminativity: In some systems, for every accent domain there is a single major prominence peak. This does not preclude the possibility that other syllables in the same domain may also be prominent relative to surrounding syllables (see further below under RECURSIVITY), but there is only one which is the most prominent one of them all. In a non-culminative system, there may be two prominences within the same domain without either of them being more prominent than the other one (in some languages, e.g. Chinese, accentuation in compounds appears to be non-culminative).

It is a matter of debate whether it is useful to distinguish obligatoriness and culminativity. The alternative is to operate with a single parameter, usually also called simply culminativity, defined as the property where every lexical accent domain has a single major accentuation. If one separates culminativity (in a narrow sense) and obligatoriness, languages such as Japanese have a non-obligatory, but culminative accent-system (i.e. not every word has to have an accent, but those that have an accent have only one). If one operates with a single parameter culminative (in a broad sense), then Japanese is non-culminative since not every word has an accent.

Recursivity: In some languages, it is possible and useful to distinguish different levels of lexical accentuation. Thus for English, for example, one commonly distinguishes at least three different levels of syllable prominence: primary accent, secondary accent, and unaccented. Primary accent is assigned to the most prominent syllable in a word (as the English accent system is culminative, there can be only one such syllable). Secondary accents are assigned to syllables which are also somewhat prominent and in certain contexts can actually become carriers for the primary accent. There can be several of these in an English word, as in *èxtramètricálicity* (using grave accents to mark secondary accents). However, in some languages there is no evidence – or at best very weak evidence – for anything resembling secondary accent: a single accent is assigned to a word domain, and all the other syllables are simply “unaccented”.

One widely-adopted analysis of such secondary accents in languages that have them is in terms of sub-word domains called (*metrical*) *feet*. In a word with secondary accent, the word domain consists of two or more feet, each with its own most prominent syllable, and one foot is singled out as the most prominent foot of the word. The prominent syllable of the prominent foot is the primary accent; the prominent syllables of the other feet are secondary accents. In languages without secondary accent, we may say either that there is no level of structure corresponding to the foot, or that the feet are “unbounded”, i.e. that they are coextensive with the word. See Ewen & van der Hulst (2001) for a comprehensive introduction to metrical structure.

Transitivity: Just as accentual prominence may apply within domains smaller than the word, so we may also find accentual prominence relations at the phrasal level when words are joined together to form phrases. Within a phrase such as *yellow paper* one word (normally *paper*) is more prominent than the other word, which entails that its most prominent syllable is more prominent than the most prominent syllable of the other word. That is, the most prominent syllable of the most prominent word becomes the most prominent syllable of the phrase, often called phrasal prominence or *sentence stress*. However, not all accent systems have this feature of transitivity, and then it is not possible to single out one accented word as the most prominent in its phrase.

Phrasal prominence can be analysed in the same way as lexical secondary accent, in terms of nested domains each with its own most prominent constituent. However, not everyone accepts this point of view. In some analyses, phrasal prominence is treated as being qualitatively different from lexical prominence: on this view, lexical prominence is usually described as “stress”, and phrasal prominence is described in terms of intonational “pitch accent” (see e.g. Selkirk 1984 or Shattuck-Hufnagel and Turk 1996). For this reason it is extremely difficult to make reliable and generally acceptable typological statements about these matters.

Intonational anchoring: In many languages, as we saw in sec. 6.3.1., a lexically accented syllable serves as the ‘anchor’ for the pitch accents that make up the intonational tune. This means that in e.g. English and German the lexically most prominent syllable of the most prominent word in an utterance also carries an intonational pitch accent. This is the basis for the view of transitivity sketched in the preceding paragraph: according to this view, lexical accent is phonetically “stress”, while phrasal prominence is “pitch accent”. We prefer to see this as a fact about the relation between the accentual system and the intonational system of a given language; lexical accents may or may not serve the role of intonational anchors. In Japanese and many other languages with melodic accent, for example, there is no additional intonational feature that targets accented syllables. But this is not a function of having a melodic rather than a dynamic lexical accent: in Swedish and Basque, syllables marked with a melodic lexical accent may additionally also serve as anchors for an intonational pitch accent. Conversely, recent work on the Papuan language Kuot (Lindström & Remijsen 2005) suggests that it has dynamic lexical accent (phonetic stress) but that the intonational pitch accents do not have to occur on a stressed syllable.

Lexical distinctiveness: Finally, another commonly drawn typological distinction among accentual systems is that between fixed or predictable accent and lexically distinctive accent. In both Greek and Japanese, despite the fact that the former uses

dynamic accent and the latter melodic accent, the location of accent can be used to signal differences between one lexical item or another. In other languages, the position of stress is either completely fixed (as on the initial syllable in Hungarian or Czech) or entirely predictable (e.g. Latin, where the penultimate syllable is accented if heavy, otherwise the antepenultimate).

The dimensions of accentual typology just discussed are probably not completely independent. Accentual systems with dynamic accent (or phonetic stress) typically have obligatory and culminative lexical accent, exhibit recursivity and transitivity, and involve intonational anchoring, and in fact it is widely assumed that all dynamic accent systems exhibit these properties more or less by default. Although there is no doubt that the dynamic accent systems of Europe typically show this cluster of features, for fieldworking purposes we strongly advise you not to take this as given. Kuot and Wolof are examples of languages with phonetic stress which show that one should be prepared to encounter unusual combinations and to try to provide substantial evidence for each of the parameters.

Finally, since melodic accents are realized primarily by pitch changes, they are sometimes difficult to distinguish from tonal distinctions, and in a number of cases there is an ongoing discussion whether a given language is better analysed as a tone language or a melodic accent language. This problem typically arises when there are only two distinct pitch patterns (high/low or marked/unmarked) and when the pitch pattern changes only once per lexical item. This type of accent system is widely attested African and Papuan languages and often discussed under the heading of ‘word melody’ (see Donohue 1997, Hyman 2001, Gussenhoven 2004, for examples and discussion). The core issue in analyzing these languages is whether tonal marking has essentially a paradigmatic function, distinguishing one lexical item from the other, or rather a syntagmatic (or organizational) function, rendering the marked syllable(s) prominent in comparison to the neighbouring syllables. While this distinction is reasonably clear on the conceptual level, there are many borderline cases in actually attested systems which may be quite difficult to assign to either category (Hyman (2001) uses a set of parameters similar to the ones above for distinguishing typical tone and accentual systems). The existence of such borderline cases is not surprising given the fact that prototypical lexical tone systems may change into melodic accent systems and vice versa.

In concluding this section, a note on the ambiguity of the term “pitch accent” as used in much of the literature is in order. This term is now regularly used in two distinct ways: on the one hand, it refers to the sentence-level (intonational) pitch features that may accompany prominent syllables in an utterance in a language like English; on the other hand it refers to the word-level – lexically specified – pitch features that accompany accented syllables in a language like Japanese. In this chapter, we have opted to use the term *pitch accent* only for intonational pitch features and use *melodic accent* for lexically specified accentual pitch features.

5.4 Working on prosody in the field

In approaching the analysis of segmental phonology or morphosyntax in an unfamiliar language, there are various well-tested techniques for determining the elements and structures one is dealing with. These include things like minimal pair tests and permutation tests, which are described in more detail in Chapters 5 and 7. For certain purposes, these are also relevant for prosody – for example, we have already described the existence of lexical minimal pairs that differ only in tone, and once you have determined that you are dealing with a lexical tone language it may be both possible and appropriate to elicit minimal pairs for tone in exactly the same way that you would for segmental differences. However, to the extent that prosodic features are not organised like ordinary segmental phonological and morphosyntactic features, different techniques are required.

The most important problems in studying prosody in the field are the fact that prosody is pervasive – you can't have an utterance (even a single elicited word) without prosody – and the fact that it is influenced by both lexical and sentence-level factors and may thus be contextually variable in ways that are difficult to anticipate, or to notice. For example, if you were asked out of context to give the name of the famous park in the middle of London where people come to make speeches to anyone who happens to want to listen, you would say *Hyde Park*, with the two words about equally prominent. However, if you were in a conversation about great urban parks – like Grant Park in Chicago or Central Park in New York or Stanley Park in Vancouver – you would probably say *HYDE Park*, with the main prominence on *Hyde*. (In fact, if you read the previous sentence aloud you will find it is very difficult to say the list of park names without putting the main prominence in each on the proper name and de-emphasising *Park* in each case.) If you were doing fieldwork on English and knew nothing about the language, you would have to become aware of this contextual effect before you could accurately describe the prosody of expressions like *Eiffel Tower* or *Princes Street* or *Van Diemen's Land* that consist of a proper noun and a common noun.

In this section, therefore, we will discuss research procedures which are particularly useful in prosodic research but rarely used in working on other aspects of the grammar of a given language. We begin by describing some useful “first steps” to take in the prosodic analysis of a previously undescribed language.

5.4.1 First steps

It is important to establish early what sort of lexical prosodic features are found in the language you are working on. The literature on neighbouring and related languages may provide important pointers in this regard, but it is obviously necessary to remain open to all possibilities until clear language-internal evidence points in one direction or the other. If you are working on a language with distinctions of quantity or of lexical accent (whether dynamic accent or melodic accent), it may take some time to become aware of the distinctions, because as we noted earlier the functional load of such distinctions may be relatively low. If you are working on a prototypical lexical tone language, it is likely to become evident quite quickly, because native speakers will usually point out to you that items that you appear to consider homophonous are not homophonous but clearly distinct for them. However, unless you are working with speakers who are also familiar with a well-described tone language, they will not

necessarily make reference to tone (or pitch) in pointing out these differences. They may simply assert that the items in question sound very different, sometimes perhaps even claiming that the vowels are different.

Although there may be some languages with no lexical prosodic features whatever, in general it will be a useful starting hypothesis that in any given utterance some prosodic features will be lexically determined and some determined at the phrase or sentence level. Both levels are inextricably intertwined; there is nothing in the signal to tell you whether a given pitch movement is lexically motivated (e.g. lexical tone), intonationally motivated (e.g. sentence accent), or even both (e.g. the combinations of lexical and intonation tone commonly found on sentence-final syllables in Chinese or Thai). This problem is of central importance when analyzing pitch, but sometimes affects the analysis of quantity and accent as well. Perhaps the most important lesson to begin with is that recording and analyzing words in isolation does not in any way provide direct, untarnished access to lexical features. This is a classic mistake, unfortunately widely attested in the literature. A single word elicited in isolation is an utterance, and consequently cannot be produced without utterance-level prosodic features. For example, if you compare ordinary citation forms of the English words *PERmit* (noun) and *perMIT* (verb) [[link to sound files](#)], you might conclude that high pitch, followed by a fall, is a feature of lexical stress in English. However, high pitch associated with the stressed syllable is actually a feature of declarative statement intonation in short utterances. If you utter the same words as surprised questions, the stressed syllables will be low, followed by a rise in pitch to the end. If you utter the same words as part of a long and boring list, the stressed syllables may be high, but followed by a rise in pitch (rather than the fall seen in the isolated citation form). In short, even for single word utterances it is not a straightforward matter to distinguish between lexical and intonational prosodic features. There is no intonationally unmarked ‘citation form’; every utterance has intonation.

In order to separate the two levels, lexical items have to be observed in a number of different syntactic and semantic-pragmatic contexts. Whatever prosodic features remain constant across these contexts most likely pertain to the lexical level; features which change may relate to the sentence level. But especially in dealing with lexical tone languages, even this statement needs qualifying, because in many such languages there are complex locally-conditioned variations in tonal pattern, sometimes called *tone-sandhi* (see Yip 2002 for examples and discussion).

In order to elicit target words in different contexts, one can construct short clauses or phrases where the target words may occur in different positions (i.e. initial, medial, final). A particularly useful variant of this technique is to record short (3-5 word) lists of target words with the words in different positions in the list. If speakers produce a coherent list rather than a sequence of minimal utterances, the result is likely to be a contrast between list intonation and minimal declarative utterance intonation. This may allow you to distinguish word-level prosodic effects. More generally, list intonation may be particularly useful in the initial stages of such an analysis for three reasons. First, it is relatively easy to elicit naturally: the act of listing elicited items does not differ in principle from listing items as part of a procedural description, whereas enacting a question is quite different from actually asking a question. Second, list intonation tends to be fairly simple in the sense that there is usually only an opposition

between non-final and final members, or sometimes a three-way distinction between non-final, penultimate and final. In particular, there are no differences of information structure (focus, topic) in lists, which often complicate the interpretation of prosodic features in other types of examples (see also section 6.2.2.3 above). Third, list intonation may be more consistent across speakers, which makes it easier to recognize the same intonational targets across speakers and at the same time provide an indication of inter-speaker variability.

5.4.2 Elicitation

All modern descriptive and documentary fieldwork includes the recording of a substantial corpus of (more or less) spontaneous ‘texts’ (where ‘text’ subsumes all kinds of communicative events including conversations, narratives, oratories, etc.). If these recordings are done with reasonable quality, they can form the basis for subsequent auditory and instrumental analysis of many prosodic features of connected speech, features that may be difficult to observe in structured interview sessions and difficult for most native speakers to be aware of. However, just as you would not expect to study phonology or syntax solely on the basis of a recorded corpus, so in the case of prosody it is important to complement recorded texts with elicited data.

In eliciting data for prosodic analysis it is important to keep various factors in mind that are of only secondary importance for eliciting many other kinds of data. First and most important, it is essential to keep in mind the kind of effects that context may have, and to adjust elicitation procedures accordingly. For example, in English it is common for WH-questions to be pronounced with an overall falling contour in neutral contexts (*Where is he going?*), a relatively high level followed by a low rise at the end in polite contexts (*Where would I find Dr. Anderson?*), and an overall rising contour in repetition or reminder contexts (*Where did you say you were from?*). Eliciting such distinctions may require you to get native speakers to put themselves mentally in different contexts, which is not necessarily easy to do. We return to this topic at some length in the next section.

Second, it is important to record several speakers rather than relying on one or two primary consultants. One reason for this is the conspicuous difference of voice pitch between males and females; another is that many prosodic features vary more between individuals and between socially defined groups than do centrally “linguistic” features. Fieldwork situations will usually put severe limits on how many speakers you can work with, but if at all possible it will be valuable to record elicited material from at least 4 and as many as 8 or 10 speakers. Gender balance is an important concern in putting together a set of speakers. In situations where it is impossible to find several speakers for the same task, it may be useful to record the same material with the same speaker a few days or weeks apart. There is little use in recording the same example set twice as part of the same session because this will almost certainly produce repetition effects.

Third and finally, it is important to keep in mind that instrumental acoustic analysis is increasingly regarded as an essential part of reliable descriptions of prosody, and that preliminary instrumental work in the field may be invaluable for guiding your work. This means that elicitation must be done in such a way that the resulting recordings are usable for instrumental analysis. In devising test examples for prosodic features, it is

important to pay attention to the segmental make-up of the example in order to minimize microprosodic effects (see sec. 6.4.5). However, it is often not possible to come up with materials that perfectly control for microprosody; either the phonotactics of the language may prohibit certain sequences that would be useful to include in your materials, or the only lexical exemplars of a particular sequence may create meaningless, obscene or ridiculous sentences that native speakers may refuse to say or will be unable to say naturally. As usual in experimental work, there is a trade-off between naturalness and the control of interfering variables.

5.4.3 Problems in prompting speakers

As the example of English WH-question intonation makes clear, eliciting example sentences for prosodic research requires attention to various factors that are not usually of concern to fieldworkers, and makes demands on speakers that ordinary phonological and syntactic fieldwork may not. Suppose you carefully construct a question-answer pair, paying attention to both pragmatic plausibility and segmental make-up. It is not enough to get native speakers to produce the segments of which the example sequence consists; they have to produce the first part as a question, the second as an answer. Do not underestimate the problems involved in explaining the idea of pretending to pose a question or give an answer. Moreover, be aware that some speakers may be unable to do things like this naturally, even if they understand the idea. This is one of the reasons why it is important to record multiple speakers wherever possible: without being able to compare across a sample there is no way of knowing who is acting reasonably well and who is doing something else.

We just spoke of carefully constructing question-answer pairs for native speakers to produce, but there is a significant problem of how to present tokens for prosodic research without unduly influencing the speakers. It is of little use to have a speaker repeat what the fieldworker is saying, since there may be direct effects of repetition on the speaker's production, or the speaker may in some way imitate the researcher's model. If you are working in a literate community, reading can be a good method for eliciting intonational data, provided that the speakers understand the need to vocally enact the illocutionary force of the example sentences. Unfortunately, it often happens that even literate speakers are unable to read fluently in their native language; it is common to find speakers who are literate in a majority or national language but have little practice or experience reading their native language. One technique that has been successfully used with such speakers is to present them with material written in the language they are comfortable reading, and ask them to give equivalents in their own language. But only some speakers will produce natural-sounding utterances under such conditions. It is also known from work on major European languages that the intonation patterns found in reading do not perfectly match those found in spontaneous conversation. Here the influence of the standard norm may be a major issue.

If reading is not feasible, various role-playing and experimental tasks may be useful. For example, rather than constructing question-answer sequences in advance and asking speakers to 'enact' them as naturally as possible, one may try to involve speakers in some kind of game or role play which requires them to ask questions. A technique widely used for this purpose involves matching tasks where one speaker instructs another speaker in reconstructing an arrangement of figures, pictures or points

on a map which is only visible to the instructing speaker (see map task, space games, etc. /REF/ and chapter ??[needs input by Dan]). Another technique is to have speakers look at a picture sequence or watch video clips and then to describe these or comment on them (pear film, frog story, etc. /REF/ ?? [needs input by Dan]). The big advantage of these techniques is that speakers are prompted with non-linguistic materials, and relatively spontaneously produce naturalistic speech. Moreover, unlike completely open-ended tasks such as recounting narratives or engaging in free conversation, these tasks permit a certain degree of control over what speakers will do, which makes it possible to collect comparable data from several different speakers. While it is rare that speakers produce completely identical utterances in these circumstances, a well-devised task usually requires them to use particular words, phrases or constructions and to engage in specific linguistic routines such as asking questions or giving directions.

Such tasks are not without their problems, however. The major problem is that speakers in small and remote communities are generally not familiar with the idea of role-playing or experiment and may be unable or unwilling to participate. It is not unknown, for example, that speakers who are asked to retell a video clip they just watched comment on the colors of the main participant's clothes or the nature of the setting rather than the action depicted in the clip. Considerable time and ingenuity may thus be required in adapting the experimental set-up to the specific circumstances found in a given speech community and in explaining the task.

5.4.4 Perception experiments

For prosodic analyses it may also be desirable to obtain some perceptual data in addition to the production data generated with experimental tasks or documented in narratives and conversations. Perceptual data are needed to answer questions such as: Do native speakers actually perceive prominences at those locations where they appear in the acoustic data (or where they are perceived by the fieldworker)? Which of the various factors contributing to a given prominence (intensity, duration, vowel quality, change and height of pitch) is the one actually of major importance in the language at hand? Which parts of a pitch contour are actually perceived as major cues for question intonation? Such questions can generally only be answered with some degree of certainty by devising perceptual tests, i.e. manipulating the prosodies of example clauses or phrases and testing speakers' reactions to them. For example, one may reduce the duration of putatively stressed syllables and ask speakers to identify stressed syllables in tokens computationally modified in this way, comparing the results with results obtained when identifying stressed syllables in naturalistic (unmodified) tokens. See van Zanten et al. (2003) and Connell 2004 for detailed descriptions of such experiments.

Once again, however, it has to be pointed out that administering such experiments is not a straightforward matter and will not necessarily produce satisfactory results. Apart from problems involved in getting speakers to participate at all in a listening experiment (in some instances, putting on a headset may already be a problem), the main problem pertains to defining a task which speakers are able to perform and which also generates relevant data. In most non-literate societies, it will be impossible to use concepts such as syllable or prominence in explaining a task. Task types that may work

- to a certain degree at least - are: a) asking speakers to comment in a general way on prosodically modified examples (which produces very heterogeneous and non-specific results but may still be useful in providing pointers to relevant parameters); b) tasks which involve the comparison or ranking of similar tokens (Which of these two items sounds 'better'/'foreign'? Which token would you use when speaking to your mother? etc.).

5.4.5 Computer-aided acoustic analysis

Perception experiments of the kind just mentioned presuppose the use of programs for acoustic analysis such as Praat, Emu, Wave Surfer or Speech Analyzer (see chapter ??). Use of such programs is strongly recommended for all kinds of prosodic analyses. The main reason for using them is that they may be of help in overcoming biases in one's own perception of prosodic data and in detecting phenomena one has not been listening for. As further discussed shortly, acoustic data are always in need of interpretation and auditory crosschecking. Nevertheless, they provide the only objective source of prosodic data, and an analysis which goes against major acoustic evidence is almost certainly false.

The programs just mentioned provide fairly reliable acoustic analyses of duration, intensity and F0. These can be done on a laptop in relatively short time and hence are feasible also in field situations provided that laptops can be used at all. Handling the programs can be learned in a few hours (in particular in the case of Speech Analyzer or Wave Surfer). Hence, it would be most inefficient not to use these tools when tackling the prosodic analysis of a previously undescribed language.

The current section briefly reviews the most important parameters to keep in mind when interpreting F0 extraction. For effective fieldwork it is not necessary to understand the mathematical and engineering aspects of F0 extraction. However, it is necessary to know something about the factors that affect F0 in order to interpret pitch contour displays appropriately and to select speech materials for phonetic analysis. It is easy to be misled by what you see on the screen, and easy to make instrumental measurements that are nearly worthless.

The rate of vibration of the vocal cords can be briefly but substantially affected by supraglottal activity – that is, by the fact that specific vowels or consonants are being articulated at the same time as the vocal cords are vibrating. Such effects are often collectively referred to as microprosody. Fig. /// shows instrumental displays of two English utterances, pronounced with pitch patterns that are impressionistically the same (listen to sound files ///). However, it can be seen that the two pitch contours look very different.

a Is that Betty's kitten?

b Are you Molly Newman?

The most obvious difference is that in (b) the contour is continuous, whereas in (a) there are many interruptions. This makes sense if we recall that we must have voice to have pitch: voiceless sounds have no periodic vibration and therefore no F0. As listeners we are scarcely aware of these interruptions, but on the screen they are very

conspicuous. Even more conspicuous is the fact that the F0 in the immediate vicinity of the interruptions jumps around a lot. These so-called “obstruent perturbations” are caused in part by irregular phonation as the voicing is suspended for the duration of an obstruent, or (in the case of voiced obstruents) by changes in airflow and glottis position as the speaker maintains phonation during partial or complete supraglottal closure. Even an alveolar tap (as in *Betty*) often causes a brief local dip in F0; a glottal stop (as in *kitten*) often causes a much greater local dip. The consequence of such obstruent perturbations is often that the pitch contour on a vowel flanked by obstruents (like the first syllable of *kitten*) looks like an abrupt fall on the visual display, even though perceptually and linguistically there is no significant pitch change on the syllable at all.

Methodologically, the existence of obstruent perturbations means that great care must be taken in interpreting visual displays of F0. Beginners tend to overinterpret what they see on the screen. In case of a conflict between what you see on the screen and what you hear, trust your ears! Obstruent perturbations also mean that the best samples of speech for making instrumental measurements of pitch are stretches containing as few obstruents as possible.

The other type of microprosodic effect that it is important to be aware of is “intrinsic pitch” or “intrinsic F0” of vowels. The phenomenon here is very simply stated: vowel quality affects pitch. Other things being equal, a high vowel like [i] or [u] will have higher F0 than a low vowel like [a]. If you say *to Lima* and *a llama* using the same intonation pattern and being careful not to raise or lower your voice between the two, the F0 peak on *to Lima* will be higher than that on *a llama* even though they sound exactly the same. This effect appears to have some biomechanical basis, although it is not entirely clear what that basis is. No language has ever been discovered to be without intrinsic F0 effects, although in some languages with more than two lexically distinct level tones the effect may be smaller than in other languages.

The methodological significance of intrinsic F0 is that if you want to measure F0 level instrumentally, you need to control vowel quality. Don’t try to compare measurements of mid tones and high tones if all the mid tones occur on [i] and all the high tones occur on [o]. Be sure to compare like with like.

5.5. A final thought

In addition to being a central part of any language description, prosody is relevant to the fieldworker in a very different way, because it may affect communication with native speakers and local authorities. It has frequently been suggested that misunderstandings in cross-cultural communication can be caused by misinterpreting prosodic cues. Although there are certainly generalizations about the sentence-level uses of prosody that are valid in language after language, the details may differ in crucial ways. What sounds rude and aggressive to one party may just be the normal way of marking emphasis for the other. A noticeable fall in pitch at the end of a unit may signal a simple assertion to the non-native hearer, but the speaker actually intended to pose a polite question. And misunderstandings may occur even if the fieldworker and the community members use a contact language to communicate, because both

parties will tend to bring their native prosodic systems to the contact language. So an appreciation of the ways in which prosody can differ from language to language is in itself an essential tool for successful fieldwork.

References

Abramson 1962

Beckman, Mary E., 1986, *Stress and non-stress accent*, Dordrecht: Foris

Chafe, Wallace L., 1994, *Discourse, Consciousness, and Time*, Chicago: The University of Chicago Press

Chuang 1958

Croft, William, 1995, "Intonation units and grammatical structure", *Linguistics* 33:839-882

Cruttenden, Alan, 1997, *Intonation*, 2nd ed., Cambridge: Cambridge University Press

Donohue, Mark, 1997, "Tone system in New Guinea", *Linguistic Typology* 1:347-386

Ewen, Colin J. & Hulst, Harry van der, 2001, *The phonological structure of words*, Cambridge: Cambridge University Press

Gussenhoven, Carlos, 2004, *The phonology of tone*, Cambridge: Cambridge University Press

Hyman, Larry M., 2001, "Tone systems", in: Martin Haspelmath, Ekkehard König, Wulf Oesterreicher & Wolfgang Raible (eds), *Language Typology and Language Universals*, Berlin: de Gruyter, 1367-1380

Jun, Sun-Ah, 1998, "The Accentual Phrase in the Korean prosodic hierarchy", *Phonology* 15: 189-226

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

Lambrecht, Knud, 1994, *Information structure and sentence form : topic, focus, and the mental representations of discourse referents*, Cambridge: Cambridge Univ. Press

Laver, John. 1980. *The phonetic description of voice quality*. CUP

Lehiste, 1970

Lindström, Eva & Bert Remijsen, 2005, "Aspects of the prosody of Kuot, a language where intonation ignores stress", *Linguistics* 43(4), 839-870

Pike, Kenneth L., 19

Rialland, Annie & Stéphane Robert, 2001, "The intonational system of Wolof", *Linguistics* 39: 893-939

Selkirk, Elisabeth O., 1984, *Phonology and Syntax: The relation between sound and structure*. Cambridge, Mass.: The MIT Press

Sluijter and van Heuven 1996

Steedman, Mark, 2000, *The syntactic process*, Cambridge, MA: The MIT Press.

Van Zanten, Ellen, Robert W.N. Goedemans and J.J. Pacilly, 2003, "The status of word stress in Indonesian", in: Van de Weijer, Jeroen M., Vincent J.J.P. van Heuven and Harry G. van der Hulst (eds), 2003, *The Phonological Spectrum II: Suprasegmental Structure*, Amsterdam: Benjamins, 151-175.

Yip, Moira, 2002, *Tone*, Cambridge: Cambridge University Press