Comp. by: PG2846   Stage : Revises2   ChapterID: 0002012552   Date:20/11/13   Time:13:46:35
Filepath:d:/womat-filecopy/0002012552.3D493

OUP UNCORRECTED PROOF – REVISES, 20/11/2013, SPi

16

Methodology of studying intonation: from data collection to data analysis

SUN-AH JUN AND JANET FLETCHER

16.1 Introduction

This chapter describes the methodology of studying intonation, from data design and collection to data analysis, within the Autosegmental-Metrical model of intonational phonology (Bruce 1977; Pierrehumbert 1980; Beckman & Pierrehumbert 1986; Ladd 1996/2008). Methods of data design and collection would differ depending on various factors: whether the language is well-studied, under-studied, or endangered; whether the research team is familiar with the language and has an easy access to native speakers or not; or whether the data collection occurs in the field or in the lab. In this chapter, the methods of collecting intonation data are provided for two different situations, i.e. when the data are collected in the laboratory vs. in the field. The former case is typical of when the grammar of the language is well-known and the research team is familiar with the language, often having easy access to native speakers of the language and good recording facilities or environment. The latter case is typical of when the grammar of the language is under-studied or the language is endangered, the research team is not very familiar with the language in general, and the recording environment is not always optimal. In sections 16.2 and 16.3, methods of data collection are described for these two cases, respectively, and section 16.4 provides ways to analyze intonation data including criteria for deciding intonational categories. Section 16.5 concludes the chapter.

16.2 Designing and collecting intonation data in the laboratory

This section describes the basic steps of designing a corpus of data collected in the lab, when one wants to study intonation for the first time in the framework of the Autosegmental-Metrical model of intonational phonology. For this scenario, we assume that the grammar of the language is well-known and well-studied, with a

Comp. by: PG2846   Stage : Revises2   ChapterID: 0002012552   Date:20/11/13   Time:13:46:35
Filepath:d:/womat-filecopy/0002012552.3D494

OUP UNCORRECTED PROOF – REVISES, 20/11/2013, SPi

494   *Sun-Ah Jun and Janet Fletcher*

fair amount of literature describing the word prosody, i.e. whether the language has a stress or tonal specification at the lexical level, though the status of word prosody is often controversial. Knowledge of the lexical property of a word is important in studying intonation because the intonational contour of a sentence is composed of tonal properties associated with the word as well as of the phrase. (For the same reason, studying word prosody would not be possible without considering intonation.) Therefore, the first subsection (section 16.2.1) describes ways to discover basic intonational categories (e.g. pitch accents, phrasal tones, boundary tones) of a language in declarative sentences, given various types of word prosody. If the word prosody of a language is controversial or unknown, readers should try to use all methods suggested in this section. Section 16.2.2 and section 16.2.3 describe how to investigate intonation patterns in other sentence types and focus prosody, respectively, and section 16.2.4 describes how to investigate the existence of higher prosodic units in the language. Finally, section 16.2.5 provides tips on finding speakers, preparing a script, and recording the data.

To tease apart the word prosody from phrasal prosody, we need to begin the study of intonation by examining the Fo contour of a multi-word phrase. Intonation of a single-word phrase, i.e. a word in isolation, should be only examined after figuring out the basic intonation pattern of a multi-word phrase. In order to create a smooth pitch track of an utterance, it is recommended to use a sentence that consists of as many sonorant sounds as possible when first analyzing intonation. But once the basic tonal pattern is established, a sentence should include any type of sounds used in the language so that the analysis of intonation is not limited to sonorant sounds. This is crucial, as intonational tones can change depending on the segment type as in Korean (Jun 1993, 1996b, 1998).

16.2.1 *Finding intonational categories*

16.2.1.1 *If a language has stress*   If a language has stress, it is very common that a stressed syllable is realized with non-default pitch (e.g. high, low, falling, or rising). To investigate if a certain pitch target or movement on or near a stressed syllable is associated with the stressed syllable, i.e. if the language has a pitch accent, we should examine if the tonal pattern around the stressed syllable remains the same when the location of stress changes. To do this, we can vary the location of stress in a word by constructing three-word declarative sentences where the target word is three or four syllables long. (1) shows an example based on a three-syllable target word, which is placed in initial, medial, and final positions of a sentence; here "s" refers to a syllable and "S" refers to a stressed syllable. We should vary the location of stress in the target word while keeping the other two words constant so that any change in the intonation contour is likely due to the change of stress location in the target word.

It is possible that the tonal shape of the pitch accent is different depending on its location. If the pitch accented syllable is near another pitch accented syllable, the underlying tonal target of the pitch accent may not be fully realized, and if it is near the edge of a word or a phrase, it may interact with the boundary tone or the tone itself may be the head of the prosodic unit (i.e. nuclear pitch accent). For this reason, it is important to locate the target word in various positions of the sentence and inform your speaker not to put a break or pause between words when producing sentences and not to focus or emphasize any word (see section 16.2.5 for information about a speaker and data collection procedure). A word produced in isolation should be interpreted carefully because it may be produced with some degree of emphasis, thus showing focus prosody, in addition to the combination of word-level tones and phrase-level tones. When a stressed syllable is the last syllable of a word or a phrase, interfering with the boundary tone of a prosodic unit, languages differ in how they resolve the tonal crowding conflict. A syllable can be lengthened to accommodate multiple tonal categories (e.g. English, Beckman & Pierrehumbert 1986; Grabe 1998), some of the tones can be truncated (e.g. German, Grabe 1998), or the tones of the lower prosodic unit can be overridden by those of a higher prosodic unit (e.g. Korean, Jun 1993; French, Jun & Fougeron 2002; Bengali, Khan, this volume).

(1)   Varying the location of stress in each word in a three-word declarative sentence
       (s = syllable, S = stressed syllable)

	(a)			(b)			(c)		
	Sss sSs sSs			sSs **Sss** sSs			sSs sSs **Sss**		
	sSs sSs sSs			sSs **sSs** sSs			sSs sSs **sSs**		
	ssS sSs sSs			sSs **ssS** sSs			sSs sSs **ssS**		

Once we know that the language has a pitch accent, we want to know the underlying tonal target of the pitch accent and its various phonetic realizations. For this, we can increase the number of unstressed syllables before and after the stressed syllable, as shown in (2). This will show how a pitch accent is realized when stressed syllables are next to each other or when a sentence begins with a stressed syllable vs. when a stressed syllable is sufficiently far from the edge of a prosodic unit or other stressed syllables. When making up target sentences to increase the inter-stress interval, a sequence of unstressed syllables can come from one or more words. For example, both $(Sssss)_w(Sss)_w$ or $(sSss)_w(ssS)_w$ can be an example of inter-stress interval of four unstressed syllables. But this would work only if the language has no boundary tone marking the edge of a word. If the language has one, we need to put a stressed syllable away from the edge of a word to avoid any possible influence from the boundary tone on the realization of pitch accent. In sum, we can find the underlying tonal category of a pitch accent and its realizations by moving the location of a stressed syllable relative to other stressed syllables and the boundary of a prosodic unit.

(2)   Increasing the number of unstressed syllables ("s") before and after a stressed
         syllable ("S")

(a) **S**s ssSs sSs	(b) ss**S** Sss sSs		
s**S**s ssSs sSs	s**S**s Sss sSs		
ss**S**s ssSs sSs	s**Sss** Sss sSs	or	s**Ss s S**s sSs
sss**S**s ssSs sSs	s**Ssss** Sss sSs	or	s**Sss sS**s sSs
ssss**S**s ssSs sSs	s**Sssss** Sss sSs	or	s**Sss ssS** sSs

When a language has free or variable stress locations, it often has multiple types of
pitch accent. Different pitch accent types often convey different semantic and
pragmatic meanings (e.g. West Germanic languages). For this type of language, it
is important to examine the pitch patterns of stressed syllables when sentences are
produced in various semantic and pragmatic contexts.

When a language has a fixed location of stress toward the edge of a word, the pitch
accent associated with the stressed syllable functions as an edge marker of a word or a
small prosodic unit such as an Accentual Phrase (AP) (e.g. Kiche, Nielsen 2005).
Another common phenomenon of this type of language is that in addition to a pitch
accent marking one edge of a word/AP, they have a boundary tone marking the other
edge of the prosodic unit. For example, in Bangladeshi Bengali (Khan, this volume),
Georgian (Vicenik & Jun, this volume), and Tamil (Keane, this volume), a Low pitch
accent (L*) marks the beginning of an AP while a High AP boundary tone (i.e. Ha)
marks the end of an AP. For this reason, an AP often shows a tonal melody such as
rising or falling or rising-falling over the entire Accentual Phrase. An AP is slightly
larger than a word (thus most APs have only one word) and is typically marked by a
tone at the edge(s) of the phrase. If a tone is an edge tone, it would stay around the
edge of a word or phrase regardless of the location of stress; see Tamil and Bengali
chapters in this volume for data supporting the edge tone analysis.

16.2.1.2 *If a language has lexical pitch specification*   If a language has a lexical pitch
accent, it is expected that a certain syllable or mora of a word is tonally marked in the
lexicon, regardless of whether the tonally specified syllable is stressed (e.g. Swedish)
or not (e.g. Japanese). That is, the tonal pattern is word- or morpheme-specific, so it
is likely to stay the same when the discourse context or a sentence type changes. Since
Fo changes over a word are distinctive, there is presumably not much room for Fo
variation at the postlexical level. Fo changes at the phrasal level can be due to the
interaction with adjacent pitch accents, or a boundary tone marking the edge of a
prosodic unit, or a phrasal tone marking some type of prominence. To find out if Fo
changes are due to the first two cases, we can prepare sentences similar to those in (1)
and (2). That is, we can vary the location of pitch accent relative to another pitch
accent or a word boundary so that there are lexically unmarked syllables/moras
between two pitch accents and around a word boundary. The underlying tonal

targets would be fully realized when a pitch accent is sufficiently far from other tonal targets, but may be undershot or not realized when the tonal targets are near each other. If there is an Fo peak or valley at or near every word boundary, that tone is most likely a word boundary tone, but if most, but not every, word boundary shows a similar pattern of Fo peaks or valleys, those tones are more likely to be a boundary tone of an AP-like prosodic unit. Next, to find out if Fo changes at a phrasal level are due to the third case, i.e. marking prominence, we can vary the location of a word in a sentence or change the information structure of the sentence. The default tonal pattern of a word may change if the pitch accented word receives sentence focus (e.g. Swedish, Bruce 1977) or narrow focus (see section. 16.2.3 for studying focus prosody).

Furthermore, we can also vary the order of words in a sentence if the language has multiple types of pitch accent (e.g. Kyungsang Korean, Osaka Japanese) or has an accent vs. no-accent distinction (e.g. Tokyo Japanese, Lekeitio Basque). The ordering could affect the phrasing of words or realizations of the accent. For example, in Tokyo Japanese, accented words, but not unaccented words, trigger a downstep of the following accented word, and a sequence of two accented words rarely forms one AP, while a sequence of an unaccented word followed by an accented word often forms a single AP (Kubozono 1988, 1993; Jun & Koike 2008). In North Kyungsang Korean, which has multiple pitch accent types, the realization of pitch accent changes depending on the type of adjacent pitch accents (Kim 1988; Kenstowicz & Sohn 1997; Jun et al. 2006).

Finally, if a language is a lexical tone language, having lexical pitch specification on almost every syllable, it would have less room for Fo shape variation at a postlexical level than would a lexical pitch accent language. A common way to investigate if there is any intonational pitch variation in a tone language is to examine Fo contours of sentences where all syllables or all but one target syllable have the same lexical tones (e.g. all High tones or all Low tones) and see if there is any change in Fo from the underlying tone sequences when varying the sentence types or syntactic and information structures (e.g. Xu 1999; Peng et al. 2005; Schuh et al. 2010). Due to the limitation in changing tonal shapes on all or most syllables, tone languages often manipulate pitch range, duration, and intensity, or voice quality to deliver semantic and pragmatic meanings as well as sentence type information. If a tone language has lexically toneless syllables as in Mandarin, such a syllable could carry an intonational tone, and if a tone language has sentence-final particles as in Cantonese (Wong et al. 2005), the final syllable could be lengthened to carry an intonational boundary tone.

16.2.1.3 *If a language has no word prosody*   If a language has no word prosody, it is very likely that there is a phrasal tone marking the edge of each word or an AP-size unit. For this type of language, we can start with a simple three-word declarative sentence and manipulate the length of each word while keeping the other words the

same. That is, the number of syllables can be increased in the first word while keeping the number of syllables in other words the same as in (3a), then the number of syllables can be increased in the second word while keeping the other words the same as in (3b), and finally, the number of syllables can be increased in the third/last word keeping the other words the same as in (3c).

(3)   Increasing the number of syllables in each word in a sentence

	(a)	(b)	(c)
	s sss sss	sss **s** sss	sss sss **s**
	ss sss sss	sss **ss** sss	sss sss **ss**
	sss sss sss	sss **sss** sss	sss sss **sss**
	ssss sss sss	sss **ssss** sss	sss sss **ssss**
	sssss sss sss	sss **sssss** sss	sss sss **sssss**
	. . .	. . .	. . .

This will show if there is any phrasal tone forming the intonation contour of a sentence. There can be a boundary tone marking the edge of every word or a group of words (like an AP), or a phrasal tonal melody realized over one word or a phrase (e.g. Japanese accentless dialects (Igarashi, this volume), West Greenlandic (Arnhold, this volume)), or a phrasal tone realized on a specific location within a word or a phrase. If a phrasal tone is constrained by the length of a phrase, its realization could change as the phrase size changes. For example, some tones seen in a longer phrase may not be present or may get undershot when the phrase gets shorter (e.g. Korean, Jun 1998; French, Jun & Fougeron 2002). If this happens, we can tell which tone is obligatory and which tone is optional.

It is also possible that the tonal shape of a phrase changes depending on the word order (e.g. in SOV languages, a preverbal noun often receives focus, and the object and the following verb often form one prosodic unit) or the location of a word in a sentence (e.g. sentence-medial vs. sentence-final word). Often, a sentence-final word shows a different tone pattern because the last syllable of the word is likely to carry the boundary tone of the whole sentence.

To see if a sentence-medial tone is marking a word boundary or an AP type boundary, we can decrease the length of two adjacent words as shown in (4) and see if the boundary tone is still present between the two words. If the language has an AP type prosodic unit, two adjacent words will tend to form one AP if they are both short, and the tendency would increase if the two words are semantically related, if they form a tight syntactic constituent such as a noun with its attributive adjective, or if they are frequently used together (See Jun 1993, chapter 5, 1998). In (4a), the first two words are shortened while keeping the last word the same, and in (4b), the last two words are shortened while keeping the first word the same. Here, we can change the syntactic and semantic relation of the two target words and see if that affects the tonal pattern of the phrase.

(4)   Testing if multiple words can form an AP-type tonal domain

    (a) <u>**ss sss**</u> sss   (b) sss <u>**ss sss**</u>

       <u>**s sss**</u> sss   sss <u>**s sss**</u>

       <u>**ss ss**</u> sss   sss <u>**ss ss**</u>

       <u>**s ss**</u> sss   sss <u>**s ss**</u>

       <u>**s s**</u> sss   sss <u>**s s**</u>

16.2.2 *More sentence types*

After studying intonation patterns of declarative sentences, we should examine other sentence types such as interrogatives, imperatives, requests, lists, and vocatives. Interrogatives can be further divided into various types of questions, e.g. wh-questions, yes-no questions, echo-questions, incredulous questions, or rhetorical questions. Languages differ in how to mark these sentence types prosodically. They can manipulate boundary tone type, pitch accent type, phrasing, pitch range, declination slope, amplitude, duration, or some combination of these. A boundary tone is typically realized on the last syllable of a phrase, though it is possible to span over the last few syllables of a phrase. Futhermore, languages can use a boundary tone to distinguish different sentence types without affecting phrase-internal tones and phrasing.

To find out sentence-type-specific prosody, we can try to make minimally paired sentences as in (5) and compare the intonation patterns across sentence types. By keeping the same words and structures across different sentence types as much as possible, we can minimize any tonal difference associated with specific words and structures. We should also examine longer sentences and complex syntactic structures in each sentence type to see if any intonational category is repeated within each sentence type or if there are common or different features across sentence types as the length or structure of a sentence changes. For the wh-questions, we should also try sentences that include various wh-question words, e.g. *what, where, when*, and a noun phrase modified by a wh-word, e.g. *which* Noun, *whose* Noun, or *what* Noun.

(5)   Declarative: *John opened the door.*
       Wh-question: *Who opened the door?*
       Yes-no question: *Did John open the door?*
       Imperative: *Open the door!*
       Request: *Would you mind opening the door, please?*

16.2.3 *Focus prosody*

So far, we have described how to design data that do not include any focused items. Though languages differ whether and how focus is cued phonetically, intonation often plays a major role in cueing focus across languages. A focused word or a

focused constituent of a sentence is often distinguished from the non-focused part of the same sentence, or from the same word in a corresponding neutral sentence, by means of higher intensity, longer duration, and larger pitch range. But they can also employ different types of pitch accent, phrasing, or boundary tones. Though there are various ways to highlight a certain word or a constituent semantically and pragmatically, the most common types of focus that have been described prosodically have been contrastive, especially corrective, focus (e.g. It's X, not Y) and information focus (also known as question-answer focus or presentational focus, e.g. *What did you see today*? *I saw X today*; Kiss 1998; Gussenhoven 2004b; Katz & Selkirk 2011). Eliciting these types of focus is easier if we use a dialogue where the participants exchange a question and answer pair as shown in (6). It is important to add a "neutral" sentence corresponding to the focus sentence (i.e. segmentally identical to the focus sentence, but without focusing any item pragmatically) so that any intonation pattern of the focus sentence that is different from that of the neutral sentence can be interpreted as focus prosody.

(6)   (a) Neutral focus data (produced out-of-the-blue):
        John bought a book.

    (b) A short dialogue to elicit Information Focus on a single word or a phrase
        A: *Who bought a book?*
        B: *JOHN bought a book.*
        A: *What did John buy?*
        B: *John bought a BOOK.*
        A: *What did John do with a book?*
        B: *John BOUGHT a book.*
        A: *What did John do?*
        B: *John BOUGHT a BOOK.*

    (c) A short dialogue to elicit Corrective Focus on a single word or a phrase
        A: *Did Mark buy a book?*
        B: *(No,) JOHN bought a book.*
        A: *Did John buy a pen?*
        B: *(No,) John bought a BOOK.*
        A: *Did John sell a book?*
        B: *(No,) John BOUGHT a book.*
        A: *Did John watch a movie?*
        B: *(No,) John BOUGHT a BOOK.*

As shown in (6b, 6c), it is important to create a question so that the answer is the same as the neutral version but only differs from it in the presence of focus. For the (6c) type dialogue, it is helpful to print the word "No" in parenthesis to trigger the correction meaning but the negation word should not be produced out loud because it may affect the realization of the following focus sentence, which will then prevent comparing the focus sentence with the neutral sentence. The focus sentence without

"No" and the corresponding neutral sentence can be compared in detail phonetically (e.g. duration, intensity, and F0, examined on the focused word as well as the pre-focus and post-focus strings) as well as phonologically (e.g. pitch accent type or phrasal tone, boundary tone, phrasing). In general, a focused word is marked phonetically by longer duration, higher F0 peak, larger pitch range, higher amplitude, and also by stronger articulation. The post-focus string is often marked by low F0, shorter duration, and weak amplitude (e.g. Cooper et al. 1985; Eady & Cooper 1986; Xu & Xu 2005; Breen et al. 2010). The pre-focus string is also often reduced phonetically, though not as substantially as the post-focus string (Jun & Lee 1998; Baltazani & Jun 1999). Phonologically, a focused word can be marked by different pitch accent type or phrasing, and post-focus words are often deaccented (deletion of pitch accent) or dephrased (deletion of a phrase boundary) (Beckman & Pierrehumbert 1986; Ladd 1996/2008; Ueyama & Jun 1998; Jun & Lee 1998; Gussenhoven 2004b; Jun 2011).

A question-answer dialogue such as the one presented in (6) would be the simplest way to trigger focus. But not all speakers are good at imagining the focus context based on the question-answer pair only. In order to help elicit focus more naturally and easily, a more elaborate context, in a text or in action, could be given before the question sentence.

In addition to focus, intonation is often used to mark other information structure such as topic and discourse-related meanings such as new or given information (e.g. Steedman 2000; Büring 2007; Hirschberg & Pierrehumbert 1986). To study how intonation marks such semantic and pragmatic information, a richer context and background information should be provided.

16.2.4 *Investigating a prosodic structure and extending basic intonation patterns*

A prosodic structure is a hierarchical structure of prosodic units. In the AM model of intonational phonology, the highest prosodic unit marked by intonation is an Intonational Phrase (IP). Though the lowest prosodic unit assumed is a mora or a syllable in the hierarchy, the lowest unit that can be defined by intonation is a Prosodic Word (PW). Across languages that have been analyzed so far, an IP is defined by a boundary tone at its right edge (and sometimes also at its left edge), phrase-final lengthening, and an optional pause following the unit. Prosodic units that are higher than a Word and lower than an IP are an Intermediate Phrase (ip) and an Accentual Phrase (AP). These two prosodic units are sometimes called a Major phrase and a Minor Phrase, respectively (Kubozono 1988, 1993; Selkirk & Tateishi 1988).[1] As mentioned earlier, an AP is slightly larger than a Word, and is marked by an edge

---

[1] A Phonological Phrase (PhP), which is indirectly defined by syntax (e.g. Nespor & Vogel 1986; Selkirk 1986; Hayes 1989), has been claimed to correspond to an ip (e.g. English, Shattuck-Hufnagel & Turk 1996) or an AP (e.g. Korean, Jun 1998). Then, recently, Itô and Mester (2010) proposed that an ip (or a Major Phrase) and an AP (or a Minor Phrase) are two different levels of PhP.

tone or a tonal melody, and if a language has a pitch accent, an AP can have at most one pitch accent. For duration, it can have minor or no phrase-final lengthening. An ip typically includes a few words or APs and is the domain of pitch reset, though not always marked by a boundary tone, and has a medium degree of phrase-final lengthening (i.e. weaker than IP-final lengthening). All languages would have an IP and a PW, but languages differ whether they have an ip, an AP, both, or neither, and how each unit is realized.

A typical three-word sentence produced in a neutral focus context would generally form one Intonational Phrase unless it is produced very slowly. To investigate whether a language has an ip, we can increase the "weight" of a sentence, either phonologically or syntactically. To increase phonological weight, we can increase the length of each word or a phrase, and to increase syntactic weight, we can make a noun phrase heavy by including a noun modified by a relative clause, a prepositional phrase, an adjective phrase, etc. We can also examine sentences that have a coordinate structure, a subordinate clause, or a parenthetical phrase. Though prosodic structure is not isomorphic to syntactic structure, it is common across languages that large prosodic boundaries occur at the edge of sentence-internal clauses or "heavy" syntactic constituents.

Another way to investigate the presence of an ip is to make a sentence or a phrase syntactically ambiguous and ask speakers to disambiguate the structure. For example, a phrase, *old men and women*, could be parsed as (*old men*)(*and women*) or (*old*)(*men and women*), and a common way to distinguish these two meanings in English is to add an ip boundary between the two groupings (Lehiste 1973; Beckman & Pierrehumbert 1986; Price et al. 1991). Similarly, when a head noun of a relative clause (RC) is a complex NP (i.e. NP1 *of* NP2) and either of the nouns can be modified by the RC, a common way to distinguish these two structures is to add an ip boundary after NP1 or NP2 (Fodor 1998, 2002; Bergmann, Armstrong, & Maday 2008; Jun 2010; Jun & Bishop 2013).

Intonation can also mark a change in word order (e.g. Georgian, Skopeteas et al. 2009; Vicenik & Jun, this volume; Fletcher, this volume) or function as a morpheme (e.g. the absolutive case marker in Samoan, Yu 2009). Thus, studying intonation should include observing the interface between syntax and prosody. After examining the intonation of the typical word order in each sentence type, we should examine the intonation of atypical word orders and also vary and extend the complexity of morphological structure. It is also possible that a complex morphological word or a sentence with a long sequence of clitics shows atypical intonation patterns (e.g. French, Jun & Fougeron 2002). In addition to marking morphosyntactic information, intonationally defined prosodic units can also function as the domain of segmental phonological processes. So, it is advised to examine how segments are realized depending on their location within or across prosodic units. The distribution of allophones can sometimes cue the domain of prosodic units (e.g. Indian Bengali,

Hayes & Lahiri 1991a; Korean, Jun 1998; Greek, Arvaniti & Baltazani 2005, Baltazani 2006; but see Frota 2000).

Finally, we should try to analyze various types of semi-spontaneous speech (e.g. news, interviews, story readings) using the intonational categories and prosodic units hypothesized from models of "lab" speech and generalize the model of intonation. This is what has been done to evaluate the success of labeler agreement for various ToBI (Tones and Break Indices) systems (e.g. English ToBI in Pitrelli et al. 1994; German ToBI in Grice, Baumann, & Benzmüller 2005; Japanese ToBI in Venditti 2005; Korean ToBI in Jun 2000; Jun et al. 2000).

16.2.5 *Finding speakers, preparing for a script, and recording in the lab*

Intonation varies more than segmental properties across speakers. Intonational features of the same sentence type in a language can vary by dialect, age, gender, context, and degree of emphasis. So, it is important to collect intonation data from multiple speakers, both several males and several females of a similar age group, and speaking the same dialect, to find common properties of that language variety's intonation. However, it is practically easier to start examining utterances produced by a female speaker who has a wide pitch range. Utterances produced in narrow pitch range may not clearly show the contrast between Fo peaks and valleys.

We should also make sure to avoid situations that could trigger unintended focusing, phrasing, or boundary tones. For example, if we want to record a list of sentences to test certain hypotheses, target sentences should be randomized and mixed with fillers so that sentences that are minimally different are not listed next to each other or displayed on a computer screen one after another. Adjacent sentences that differ by one word are likely to trigger focus. A long sentence should not be written in two lines because speakers are likely to pause or put a prosodic break at the end of a line, not necessarily reflecting a natural prosodic grouping. A single sentence or a phrase should be written in one card or screen to prevent speakers from using a listing intonation at the end of each sentence or a phrase. It is helpful to ask speakers to pause briefly after reading each sentence.

Reading sentences one by one, out of the blue, is a common way to study a default pattern of intonation. But this will create an intonational model of the "reading" style only, which is fine if the goal is to study reading intonation, but not good if the goal is to study the intonation of more natural speech. This is because reading intonation and conversational intonation, though often sharing the same basic properties, are not necessarily the same. Conversational intonation typically has a richer inventory of boundary tones and pitch accents and has more variation in phrasing than reading intonation. To collect conversational intonation data from a list of sentences, we can modify the sentences so that they include morphemes or words that are used in conversational speech, and ask speakers to produce them as if they talk to their family

members or friends. This works fairly well unless the speaker is very reserved or too tense and nervous about their voice being recorded.

To collect intonation of more natural speech while still including controlled material to test a specific set of hypotheses, creating a dialogue between two native speakers is very useful. Making two speakers act out a dialogue helps them to engage in the conversation more easily. Embedding target sentences in each role's dialogue and switching the role between the two speakers will allow us to collect all target data from each speaker. This method is especially useful when collecting intonation of non-standard or otherwise stigmatized dialects. In that case, creating the dialogue by adding dialect-specific lexical items or morphological markers would help speakers to code-switch to their native dialect.

Using a dialogue is also useful to study focus prosody. Eliciting focus prosody is one of the hardest tasks when collecting intonation data, especially when the data are a list of sentences. To help speakers to produce focus naturally, a more detailed context should be given before starting the dialogue. Though the context and the dialogue can be tailored to trigger focus on a certain item or phrase, not every speaker is good at remembering the context or imagining the situation to produce focus as intended when they are aware of the fact that their speech is being recorded. To help speakers remember which item is emphasized, especially when the item is narrowly focused (e.g. corrective focus), a focused item can be written in a bigger font or made more visible by changing the font style (e.g. underlined, in capitals, in bold, or in italics). Speakers are generally familiar with the pragmatic meaning of different font styles, as they are often used in cartoons and other popular literature. However, changing the font style is not appropriate if the focus type is not narrow focus or if we want to study prosodic features of new vs. given information. To study prosodic features marking discourse structure or the information status of items in the discourse, we can examine the intonation of a story or passage that we create or modify to test our hypotheses. Reading a story or a passage in a lively way or acting out a dialogue with emotion would allow us to control the material and to compare intonational data across speakers.

However, though data from story reading, dialogues, or news can provide basic intonational patterns, this type of data is still scripted and may have some limitations in terms of naturalness. To increase naturalness, researchers have used close-to-spontaneous speech data from a map task (e.g. Grice et al. 1995; Stirling et al. 2001) or a game designed to examine certain prosodic phenomena (e.g. Schafer et al. 2005; Speer et al. 2011). Analyzing fundamental elements of intonation based on Fo shapes and timing, and understanding their linguistic functions could be done, to a certain level, based on lab speech (i.e. less spontaneous but more controlled speech). But in order to fully understand the form and meaning of intonation in a rich discourse context, fully spontaneous speech should be examined. However, we suggest examining spontaneous speech at a later stage, not because it is less important but because

intonation of spontaneous speech is much harder to interpret than that of controlled lab speech. Without the knowledge gained from the controlled lab data, we would not know for sure if a certain F0 change in spontaneous speech is due to a segmental interaction, syntactic structure, semantic or pragmatic context, or other factors. It would require us to examine far more data to distinguish these possibilities.

16.3 Collecting intonation data in the field

16.3.1 *Data collection*

The points outlined in section 16.2 are also highly relevant for the intonational fieldworker. In order to undertake research on intonation and intonational variation in a traditional fieldwork setting, it is crucial to have a clear idea of the corpus or corpora you may wish to record (see also Himmelmann 2006; Himmelmann & Ladd 2008 for a good general overview of prosodic fieldwork, Maddieson 2001; Gordon 2003; Ladefoged 2003 for summary of general phonetic fieldwork). For many researchers, it may be the first time they have worked with language consultants outside of the phonetics laboratory or a formal academic setting. Before embarking on any kind of phonetic or linguistic fieldwork exercise, it is important to do as much background research as possible on the language or language variety in question. This may involve looking at descriptive grammars, dictionaries, or any other archival material including existing recordings for a particular language. In the case of a less well-studied language, this task may prove challenging. It is very difficult to do intonational fieldwork on a language that has not had at least some basic linguistic or phonological description. Even when the description includes basic phonological statements about the phoneme inventory, phonotactics, and so forth, there is often relatively little included on intonation and post-lexical prosody in general.

It has been pointed out elsewhere (e.g. Himmelmann & Ladd 2008) that traditional approaches to gathering material in order to write a grammar or examine a particular grammatical feature of a language or languages include elicitation of particular forms, recording citation lists (e.g. a Swadesh wordlist, Swadesh 1952), narratives, or texts. Researchers often work with a language consultant to translate these materials soon after, or at the time of recording. Interactive discourse is often recorded but rarely analyzed, with some notable exceptions (e.g. Garde 2006 for Kunwinjku). Likewise, traditional phonetic fieldwork often focuses on the examination of specific segmental phonetic contrasts within isolated words or within words placed in static sentence frames (see Gordon 2003 for a discussion of this). Fieldwork studies of "lexical tone" can provide useful information on how to go about collecting a corpus, with the ultimate aim of an eventual analysis. For example, Hyman (2010) identifies three basic steps in the analysis of lexical tone: a) the identification of surface categories, b) the examination of contextual variation, and c) a phonological analysis that accounts

for variation observed in a) and b). Lists of minimal pairs, and where possible, suitable sentence frames of different lengths or structure are constructed to control for potential interactions with other aspects of the grammar (after Pike 1948). Tokens of interest can also be placed in different sentence positions to ensure they are in discourse-prominent contexts (e.g. see Remijsen's 2007 study of Magey Matbat).

Given the notoriously slippery nature of the form/function relationship with intonation, it is always possible to replicate paradigms that are conventionally used to investigate segmental or lexical tone contrasts. So where does the analyst start with an intonation analysis? Some would argue that it is still important when working on a new and perhaps previously undescribed language, to start with lexical prosody. However, it is important to take into account the complex relationship between word and phrase-level pitch patterns, so for this reason, a better approach is to follow the one outlined in 16.2.1. This is feasible if the analyst has sufficient knowledge of the language or access to materials, or can work with a language consultant or linguist who can assist with the task of material development. In the case of morphologically polysynthetic languages, it should also be possible to start with a less complex morphological sequence of two words, for example, and then add words or word elements to increase the complexity of the utterance.

At an intonation workshop some years ago (WaveIP 2000), a group of investigators discussed how to go about constructing an intonational analysis from scratch. A key conclusion was that while it is important to consider what labels or categories are necessary within a particular prosodic framework (e.g. an autosegmental framework), one of the primary goals should be to discover what the significant categories are for the variety or language in question. A study of a new language will also require careful consideration of the relevant non-prosodic categories of the grammar that will include syntactic structure, sentence modality, focus-marking strategies, and information structure (e.g. Fitzpatrick 2000; Beckman & Venditti 2011). It is also important to bear in mind that intonation may not figure in focus-marking strategies if the grammar uses particles or other structural features to signal focus as in Navajo, for example (McDonough 2002).

With regard to fieldwork languages, it is likely that the analyst is not a speaker of the language; thus, as mentioned earlier, they should ideally work with language consultants as well as a linguist who is familiar enough with the language that they can understand how to formulate relevant questions for the study. These days, such questions are less informed by so-called "universalist approaches" to form/function relationships than by more detailed questions on the phonetics and phonology of intonation patterns within a particular language (see Ladd 1996/2008, and Arvaniti 2012, for a discussion). There is also a general interest in working out ultimately where the language might fit in relation to some kind of intonational or prosodic typology. Recent studies providing overviews of prosodic and intonational typology (e.g. Ladd 1996/2008; Jun 2005b; Beckman & Venditti 2011) can clearly help a

researcher to describe and model the post-lexical (and also possibly lexical) pitch features of a language.

A complication for intonational fieldwork is that there are relatively few complete descriptions of word-level and phrase-level prosody of the world's languages, meaning our existing typologies are limited. Nor should we assume that assigning a language to a typology might be a simple process of ticking off a list of parameters or features that are typical of one type of language compared to another (i.e. "pitch accent" vs. "stress accent" and so on). Many languages fall between traditional typological groupings or show features that would put them in more than one conventional category (e.g. Mandarin has both lexical stress and contrastive tone). One may also have to contend with preconceptions about the prosody of the language or language family that have been established for many years. For example, the conventional assumption within Australian linguistic circles is that all indigenous Australian languages are "stress languages" (e.g. Dixon 1980). The main quantitative studies of post-lexical accentuation however, show there is very little evidence of culminative prominence in the classic sense of the term (see Pentland 2004 for an analysis of prominence in Warlpiri, and also Bishop 2003 for an excellent discussion of this in the context of Australian languages in general). Also, it cannot be assumed that all languages within a language family are going to have identical prosody and intonation. There is an emerging view within the autosegmental framework that "tone" can come from either a lexicon of "pragmatic morphemes" and/or a set of lexical contrasts akin to vowel and consonant contrasts in segmental analysis (Beckman and Venditti 2011: 38–9). It is further suggested that a more nuanced multi-valued typology is perhaps more useful, particularly in studies of less-well described languages (e.g. see the discussion in Hualde 2006; Gooden et al. 2009).

A great deal of descriptive intonational research of relatively well-studied languages has been directed towards exploring the relationship between intonational tune and phrasing with different types of syntactic structures or clausal relations in different types of sentences. These typically include declaratives, sentences with subordinate clauses or parenthetical clauses or co-ordinations, absolute or wh-questions, and so on (for a good example of this, see the descriptions of intonational systems included in Hirst & Di Cristo 1998). While it may be still relevant to explore questions of particular interest to older universalist studies of intonation—including the relationship between high or sustained high pitch at the end of utterances with some kind of yes/no questioning function, falling pitch with the end of declarative utterances, raised pitch range at the beginning of a new discourse segment, or elevated or expanded pitch range of an intonational constituent as a focus marking strategy, and so on—it is well-known that these patterns are not universal. For example, we now know that languages as varied as Hungarian (see Ladd 1996/2008, Arvaniti 2012), Chickasaw (Gordon 2005), Navajo (McDonough 2002), and many African languages (Rialland 2007). It may also be the case that rising questioning

tune may just be extremely rare as in the case of many Australian languages including Bininj Gun-wok (Bishop & Fletcher 2005) and Mawng (Fletcher, Singer, & Loakes 2011). In languages like Dyirbal (King 1994) interrogatives are cued mainly by a suppression of downstep within an intonational phrase and not a high rising terminal tune. In Mawng, pitch range is also manipulated, but rather than suppressing declination, the main focal constituent in an intonational phrase is "upstepped" and realized in a wide pitch range, with following material realized in reduced pitch range. Fieldworkers need to be conscious of this kind of variation (see also Beckman & Venditti (2011); Himmelmann & Ladd 2008 on this point). It is also clear that one of the pitfalls of intonational study is to assume that similar pitch patterns between two languages (e.g. one undescribed vs. another as well-researched as English) can be accounted for in exactly the same way within a particular type of theory: i.e. pitch turning points in an intonation contour are post-lexical pitch accents that may or may not be aligned with "metrically" prominent syllables in words or phrases, or that a particular kind of pitch shape signals contrastive focus compared to another more neutral pitch shape and so on. It is of course ideal if one has access to native speakers who can help determine whether particular kinds of pitch configurations are meaningful. This will only work if the analyst already has some sense of the possible range of interpretations.

Interesting problems emerge when the analyst attempts to conduct an intonational description of a language or languages with only partial grammatical description. In cases such as this, there is an assumption that there will be some degree of congruence between intonational boundaries and syntactic boundaries or some kind of discourse unit. In contemplating an analysis of a fieldwork language, there is also the question of exploring the interaction between intonation, prosodic structure, and other parts of the grammar without getting lost in theoretical circles. Ladd (2008) is very clear on this point. One strategy is to try and establish what the "default" intonational phrasing pattern might be (i.e. "explicit prosody") in a language. The next step might be to construct a varying set of materials in order to elicit this phrasing pattern. For example, Hermann (1997) examined a set of specially constructed sentences of varying syntactic structures for Balinese (e.g. nominal subjects, conjoined nominal subjects, different kinds of predicates) of varying word lengths and was able to show that intonational structure (in particular the placement of Fo peaks or pitch accents) closely mirrors syntactic structure. Moreover, as mentioned in section 16.2, the grammar of discourse prominence marking across many languages may include combinations of the following: manipulations of pitch range, intonational and prosodic phrasing, and tonal prominence, including the use of different types of pitch accents for contrastive emphasis. Languages can also deaccent material (reducing the number of pitch accents in a phrase) and/or dephrase non-focal material (reducing the number of intonational constituents) to promote a particular kind of discourse interpretation.

Ideally, the analyst should work with a variety of genres and not just narratives, or isolated utterances (see Himmelmann & Ladd 2008 for a similar range of views). As Ladd (1996/2008:281) suggests, it is a major goal of intonational research on any language to sort out what tunes occur in a language and "to be able to make explicit predictions of how a given tune will be realized when it is applied to different texts." Sometimes, in the case of severely endangered languages, the analyst might only have access to archival recordings or a set of texts that were recorded for wider grammatical analysis. There is a body of research on intonation in fieldwork languages that has been built around the analysis of narrative texts (e.g. Tuttle and Lovick (2007) for Dena'ina; King (1998) for Dyirbal; Bishop (2003) for Bininj Gun-wok) which is largely because this is perhaps the easiest kind of connected discourse to elicit, particularly when one is dealing with a largely oral tradition (as in the case of traditional indigenous communities in Australia, for example), or more simply, these might be the only source of data (in the case of Dyirbal, or other "lost" languages). While it might be the case that free narratives or stories can be associated with a particular kind of performative discourse, i.e. a type of "story-telling" intonation and prosody, it is nevertheless possible to test certain hypotheses about the relationship between final lengthening, local lowering or suppression of pitch range, and discourse segment boundaries (e.g. Tuttle & Lovick 2007), or pitch reset at the edge of new discourse segments (e.g. Fletcher & Evans 2000). A great deal of information on the interaction of intonational phrasing and syntactic groupings can also be obtained by examining a large corpus of narratives (e.g. Ross 2011; Simard 2010).

Large database projects undertaken in the 1990s on variation in several European languages sought to collect a wide variety of materials to document intonational structure. For example, the IViE corpus on English varieties in the British Isles (Grabe et al. 1998; Grabe 2004) consists of five different genres. These include context-free sentences illustrating different types of sentence modality, e.g. syntactic declaratives, polar questions, wh-questions, coordinations and the like, a fairy-tale read from a text, and a subsequent re-telling of the same story using only images. The remaining two tasks of IViE elicit interactional discourse via a map task (Anderson et al. 1991) and free conversation.

In a similar spirit, recent fieldwork intonation studies of two endangered Australian languages, Mawng and Jaminjung (e.g. Hellmuth et al. 2007; Simard 2011) used a range of elicitation tasks from the Questionnaire on Information Structure (QUIS) developed by a group of researchers based at the University of Potsdam and Humboldt University, Berlin (Skopeteas et al. 2006). These tasks were designed specifically to elicit a range of different types of monologues and interactive discourse. A similar range of materials for linguistic fieldwork is also available online from the Max Planck Institute for Psycholinguistics <http://fieldmanuals.mpi.nl>. Most of these materials can be modified to suit a particular cultural setting, and a particular research situation. They consist of a variety of materials to elicit task-oriented

Comp. by: PG2846   Stage : Revises2   ChapterID: 0002012552   Date:20/11/13   Time:13:46:37
Filepath:d:/womat-filecopy/0002012552.3D510

OUP UNCORRECTED PROOF – REVISES, 20/11/2013, SPi

dialogue, for example. One commonly encountered problem in the Australian context, at least, is that it is often difficult to record the kinds of materials that are routinely gathered in common laboratory phonology paradigms. However, Simard (2011: 34) in her examination of intonation in Jaminjung reports that while it proved difficult to record a classic experimental paradigm involving multiple repetitions of similarly structured utterances, it was quite possible to use tools that elicited more culturally relevant data including examples for dictionary entries and plant names for a number of speakers. Other work on Australian languages has also used similar sources of materials for intonational and prosodic description (Singer 2006 for Mawng; Evans et al. 2008 for Dalabon; Birch 2002 for Iwaidja; Round 2010 for Kayardild). Simard (2011) was also able to record a variety of materials using a combination of QUIS tasks, spontaneous narratives, and "stimuli-based" narratives such as the "Frog story" (Mayer 1994). Similarly, Hellmuth et al. (2007) used a combination of QUIS materials that included responses to visual materials and spontaneous narratives to explore focal pitch accent realization in Mawng.

The QUIS materials also include a specific questionnaire relating to intonation structure that seeks information about tune and sentence modality as well as questioning how phonological phrasing is realized phonetically and phonologically. The questionnaire also asks whether the language has intonational pitch accents (which of course can be very hard to differentiate from phrase edges—see Hyman 2006; Himmelmann & Ladd 2008; Beckman & Venditti 2011 with regard to these potential ambiguities), the pitch shape of the accents, whether they can be non-final, and so on. Some intonational researchers would argue that this questionnaire may not be sophisticated enough to get at the fundamentals of intonational meaning in a particular language, particularly in view of more compositional approaches (e.g. after Pierrehumbert & Hirschberg 1990, and see also Arvaniti 2012, for a useful discussion). Nevertheless, it is an encouraging development in general for a questionnaire on intonational structure to be included in a linguistic fieldwork manual of this type.

In various studies of intonation, roleplay and other interactive tasks that place the language consultant in the role of teacher can also be used (e.g. see Remijsen et al, this volume, see also Himmelmann & Ladd 2008:266 for a good overview of these tasks). This is particularly helpful if the researcher has already constructed a set of hypotheses based on preliminary impressionistic observations or on background research on pertinent linguistic features of the language in question. For example, if the researcher is exploring the interaction between accentuation and focus-marking and a particular pattern e.g. rising intonation or focal construction (early pitch peak) on a particular constituent is observed, one can repeat the pattern or patterns and ask the consultant whether the first or second is a good pronunciation, e.g. "Is this good Kunwinjku?" Similar tasks have been effective in certain fieldwork settings (e.g. Bishop 2003 for Bininj Gun-wok). In terms of roleplay, two consultants can

perform a mini-dialogue that contain specific tokens or constructions that are designed to elicit discourse prominence, and therefore potential intonational strategies that might be associated with the former. The task and subsequent analysis is made easier if focus and information structure has already been investigated in the language (e.g. Singer 2006). In all interactive tasks with a language consultant, however, it is important not to presume a level of metalinguistic awareness of traditional intonation/sentence modality relations, for example. As Himmelmann & Ladd (2008:266) point out: "Do not underestimate the problems involved in explaining the idea of pretending to pose a question or give an answer."

Another important consideration in corpus design and data recording in the field is literacy. One cannot assume that speakers will be literate, so roleplay and picture-based tasks can be extremely useful in this respect, particularly with regard to eliciting discourse-prominence marking strategies on specific tokens or constituents.

It is also perfectly possible to construct a corpus to explore a specific set of questions in a controlled way when there is sufficient information available on the language. For example, Remijsen & van Heuven (2005) examined pragmatic focus and its interaction with word-level prosody in Curaçao Papiamentu, a Caribbean Creole that has both stress and lexical tone. As mentioned in the earlier section of the paper (16.2.1), it is well known, after the pioneering work of Bruce (1977) on Swedish word accents, that it is important to consider the interaction of lexical and post-lexical pitch patterns in any exploration of word accent patterns. As Remijsen & van Heuven (2005) point out, they were able to use previously published materials on the phonetics and phonology of Papiamentu, which enabled them to formulate a set of specific hypotheses about focus and its interaction with tone and stress. Specific tokens were placed in different utterance positions; utterance final, utterance medial with orthogonal variation in focal prominence, and several repetitions recorded for a number of speakers. Similar strategies have been employed in a number of experimental studies of intonation and prosody in a group of Indonesian languages (e.g. see van Heuven & van Zanten 2007). Controlled corpora have also been recorded for a number of other fieldwork languages in order to examine specific intonational and prosodic phenomena, including among many others, Navajo (McDonough 2002), Kuot (Lindström & Remijsen 2005), Creek (Martin & Johnson 2002), Chickasaw (Gordon 2005), Dinka (Remijsen & Ladd 2008), Wolof (Rialland & Robert 2001), and Apache (Tuttle 2005), and other Athabaskan languages (see the various contributions in Hargus & Rice 2005).

16.3.2 *Finding a speaker for the fieldwork*

In situations where a language is not endangered, there are a variety of ways to find speakers of the language of interest. In some cases, the researcher already has contacts with the community via other field linguists, graduate students, or

colleagues who are working on a general grammar of a language. Many institutions also have international graduate students who may speak a minority or indigenous language in addition to the official language of their country, and it may be possible to contact speakers via the international postgraduate student organizations. Ladefoged (2003) also suggests contacting a local schoolteacher, or postmaster in a specific community. In Australian indigenous communities, this is routinely the case or it may also be possible to establish contacts through a local language center. Likewise, contact with missionaries or linguists associated with the Summer Institute of Linguistics (SIL) has been used by many researchers.

In a situation of language endangerment, finding a speaker or several speakers may prove to be more challenging as it may only be possible to make new recordings of two or three elderly speakers. In all cases, it is important to be sensitive to a range of cultural issues, with the research undertaken in an ethical manner (e.g. see Dwyer 2006 for an excellent discussion of this). Most fieldworkers would agree that it is vital to establish a good working relationship with language consultants and the community in general (e.g. Mosel 2006: 67). This could be a potential paradox for the phonetic fieldworker compared to the field linguist working on a grammar who is more likely to spend longer periods of time in the field. The current research paradigms of experimental phonetics and laboratory phonology are somewhat different from those of the typical linguistic fieldworker. Nevertheless, a lot of useful information can be gleaned from current general linguistic fieldwork practices (e.g. see Gippert et al. 2006; Thieberger, 2012).

16.3.3 *Recording and processing the data*

Recording speakers in a remote location can be difficult, particularly if a quiet room or similar location is not readily available. If recordings are to be made outside, it is best to avoid situations that are either too sunny or where there is excessive wind. As with any other type of field recording, the best quality recording devices and microphones should be used. Many fieldworkers also record high quality audio-visual data, as well as having an extra audio-recorder as backup. This can be particularly important if the analyst is also interested in a multimodal approach to intonation and prosody. There are a number of good guides available for data recording (e.g. see the Max Planck Institute website <http://fieldmanuals.mpi.nl>) , and individual websites (e.g. Bert Remijsen's website <http://www.lel.ed.ac.uk/~bert/> is a good source of useful information).

Several speakers should be recorded, ideally over several days, and materials presented in random order to avoid "order" effects. In situations of severe language endangerment, it is of course not possible to record more than two or three elderly speakers. In situations like this, Himmelmann & Ladd (2008) recommend trying to record a similar set of materials a couple of days later to avoid repetition effects that

can occur in a single recording session. When eliciting more controlled material that requires several repetitions, it will be important to mention this to the speaker before you start the recording session. For example, indigenous Aboriginal people tend to be polite in these situations but it may only be possible to record a few sets of repetitions per session per speaker. Ideally, as in any experimental study of intonation, controlled materials should try as far as possible to avoid known microprosodic perturbations (see 16.4.2) unless they are part of the primary set of research questions.

16.4 Analysis of intonation

16.4.1 *Software*

The analysis of intonation will require access to good quality acoustic phonetic software, which is freely available these days (e.g. *Praat* (Boersma 2001, Boersma & Weenink: <www.praat.org>), *EMU* (Harrington 2010), *Wavesurfer* (Sjölander 2002), and so on), in order to extract a robust Fo contour (in Hz, semitones, or ERBs) and also to allow potential analysis of other important parameters like vowel formants, intensity, duration, and phonatory effects that can contribute to prominence and edge-marking. *ELAN* (Sloetjes & Wittenburg 2008) is another useful program for analysts interested in multimodal analysis. Many of these programs also interface with each other these days (e.g. EMU and Praat, ELAN and Praat). While most intonation studies are largely based on the production of speech prosody, the tools are now readily available to perform perception studies using PSOLA synthesis routines contained in *Praat*, for example. Tune/Text relations can be tested by manipulating the alignment and scaling of tone targets, and altering pitch range, for example. It may be possible to elicit judgments of prosodic "appropriateness" of manipulated tunes, or to carry out discrimination experiments of different tunes or tune types (Vaissière 2007; Vicenik & Jun, this volume). Moreover, re-synthesis can be a useful tool if an analyst is not sure whether there is a correlation between a strong pitch movement and accentual prominence. This can be a notorious problem for researchers who come from an English-speaking background. Different alignment patterns can be synthesized by shifting an Fo peak from one syllable to another in a word (or altering other phonetic parameters if this has been observed in a production study), or even finer tonal alignment patterns can also be tested to see whether native speakers find these new utterances acceptable.

These techniques have been applied successfully in the field for the examination of lexical tone and stress (e.g. Connell (2000) for Mambila; van Zanten, Goedemans, & Pacilly (2003) for Indonesian; Goedemans & van Zanten (2007) for Javanese). Roosman (2007) also used PSOLA synthesis in her fieldwork perception study of intonational prominence in Toba Batak, a Sumatran language. She manipulated pitch height and pitch peak alignment on tokens in various utterance contexts and was able

to draw firm conclusions on the contribution of the latter to marking focal promin-
ence in Toba Batak sentences. She was also able to conclude that there is no lexical
stress in this language. The techniques used in this study were relatively straightfor-
ward and simple. Hopefully more fieldwork intonation studies will make use of re-
synthesis techniques in the future, particularly as more detailed work is undertaken
on the production of intonation and prosody in field languages.

16.4.2 *Deciding tonal categories*

Now, given F0 tracks of utterances, we need to find out what tunes exist in the
language and how to characterize the tunes in terms of distinctive tonal categories
and prosodic structure so that we can generalize and predict tunes for new texts
(Ladd 1996, 2008). That is, we need to build a model of intonational phonology of the
target language. This will require an understanding of the Autosegmental-Metrical
(AM) model of intonational phonology as well as of the effects of microprosody and
computer errors on pitch tracking. As mentioned in Ladefoged (2003, chapter 4) and
Gussenhoven (2004b, chapter 1), we should be aware of non-phonological F0 changes
such as short-term F0 changes triggered by segments, i.e. microprosody (e.g. low or
"messy" F0 points due to glottalization, high F0 at vowel onset after a voiceless
consonant, low F0 at vowel onset after a voiced consonant) and computation errors
such as pitch halving and pitch doubling. These will affect F0 contours, but these F0
changes are not phonological. Phonological F0 changes are those controlled by a
native speaker to deliver linguistic meaning or function, thus, belonging to the
intonational grammar of the language.

In the AM model of intonational phonology, an F0 contour is analyzed as a
sequence of tonal targets (High, Low, and their combinations), which can mark
either the head (pitch accent) or the edge (phrasal or boundary tone) of a prosodic
unit. A pitch accent is aligned with a stressed syllable or lexically accented syllable
while an edge tone is aligned with the edge (initial or final syllable) of a prosodic unit.
Tonal targets for pitch accents or edge tones can be identified from F0 turning points
(e.g. F0 peaks and valleys), inflection points ("elbows," e.g. an F0 point that starts
rising from a stretch of low F0), or "shoulders" (broad, plateau-like F0 peaks with a
mid or low F0 height which further falls to a lower F0). A flat F0 may need two tonal
targets, i.e. at the beginning and at the end of the plateau, before changing the
direction of F0. That is, any F0 points that are not predictable from interpolating
between tonal targets should be considered candidates of tonal target-hood (cf. a
"close copy" of the original contour (e.g. 't Hart & Collier 1975; 't Hart, Collier, &
Cohen, 1990), ignoring microprosody and F0 track errors made by software.

A tonal target should be categorical and its distinctiveness should be perceptible by
native speakers across various sentences. A tonal target is distinctive when it repre-
sents some linguistic meaning such as a question or focus, but is also distinctive when

it delivers a linguistic function such as a marker of a prosodic unit, a discourse structure, or a sociolinguistically defined group (e.g. marker of a regional dialect). That is, the "distinctiveness" can be present in any aspects of grammar. Therefore, to determine if a tonal target is distinctive in the language, we need to examine if the tonal target has a linguistic meaning or function in various phonological, morpho-syntactic, semantic, pragmatic, and sociolinguistic contexts as well as having systematic and consistent Fo shape and timing across various sentences.

The AM model of intonational phonology assumes two distinct tonal levels, High and Low, and captures various Fo values that are neither high nor low via phonetic or phonological rules (e.g. in American English, a boundary tone is upstepped after a H phrase accent and, in Japanese, a bitonal pitch accent triggers downstep). However, as the AM model was adapted to the development of ToBI (Tones and Break Indices) transcription system (Beckman & Hirschberg 1994; Beckman et al. 2005), the tonal representations and categories have become less abstract, reflecting more closely the surface Fo values (e.g. introducing a category "!H," downstep H). Furthermore, by examining more sentence types and spontaneous speech, studies on intonational phonology of various languages have shown that tonal height can be distinguished in three or four levels: High (H), Low (L), Mid (M or !H), or Super High (^H).[2] Languages differ in how many and which tonal levels they distinguish in their intonation system. For example, German and Castilian Spanish have four tonal levels (German: L, H, !H, and ^H, Grice, Baumann, & Benzmüller 2005; Castilian Spanish: L, H, M, and ¡H, Estebas-Vilaplana & Prieto 2010), Mexican Spanish has three (L, H, and M, de-la-Mota, Butragueño, & Prieto 2010) but Mongolian and Tamil have two (L, H, Karlsson, this volume; Keane, this volume). A downstep High (!H) is a type of High tone, but is realized with mid Fo because of a compressed pitch range relative to the preceding one. Thus, a downstep represents a syntagmatic relationship among High tones. Whether or not !H tone is distinctive is a language-specific property (e.g. German vs. Greek). On the other hand, a mid tone (M) is not a marker of a High tone in a reduced pitch range. It means the mid level Fo value itself is distinctive, delivering a different meaning from higher Fo (H%) or lower Fo (L%), thus showing a three-way paradigmatic contrast. So far, M tones have been proposed exclusively as a boundary tone (e.g. M%, LM% in most varieties of Spanish, see Prieto & Roseano 2010) to deliver various sentence modality or pragmatic meanings.

---

[2] So far, a Mid level tone has been found distinctive as a boundary tone only, and has been labeled as either M% or !H%. In contrast, a Super High level tone has been found distinctive both as a pitch accent and a boundary tone, and has been labeled with ^H (e.g. German ToBI) or ¡H (e.g. Spanish ToBI). In this chapter, I will use ^H for a Super High level tone simply because the "^" symbol is found on the traditional English-language keyboard, and thus easier to type.

In addition to a tonal height distinction, a tonal target can also vary in terms of its complexity. A tonal target can be composed of one (e.g. H, L), two (e.g. LH, HL), or three or more tonal targets (e.g. LHL, HLH, LHLH, HLHL). In general, singletons or bitonals are found in both head tones and edge tones, but more complex tones are typically found in the edge tones (e.g. HLH% in Bangladeshi Bengali, LHL% in Catalan and Mongolian, LHLH% in Korean), realized either on the last or the last two syllables of a large prosodic unit.

A tonal category can also represent a variation in timing, i.e. alignment to a text. A High tone can be realized at the beginning, middle, or end of a host syllable or even after the host syllable. This realization can be phonetic and does not warrant a separate tonal category (e.g. English H* followed by "<" for a delayed peak), or it can be distinctive, warranting separate labels (e.g. H* vs. H*< for "delayed high accent" in Bininj Gun-wok (Bishop & Fletcher 2005); L+H* vs. L+<H* for "delayed early rise" in Catalan (Prieto, this volume)). Similarly, a rising F0 target can rise early or late, i.e. before or after a head syllable (e.g. L+H* vs. L*+H in English). Again, this alignment difference can be phonetic or distinctive. Here, "+" implies a boundary between tone bearing units. That is, the tones before and after the plus sign are aligned to a sequence of two syllables, respectively. But if the tune–text alignment is neither consistent nor distinctive, a bitonal LH symbol without a plus sign (e.g. LH*) should be chosen to represent a rising tone (e.g. French (Jun & Fougeron 2002), Kiche (Nielsen 2005), Samoan (Yu 2009)).

Finally, we can add a diacritic to a tonal target if the tonal target has a special function. If a tonal target is a head tone marking a pitch accent, we add "*" (e.g. H*, L*+H, H*+L). If a tonal target is specific to a focused item, a diacritic "f" can be added to a focus tone (e.g. fH in Bangladeshi Bengali; Khan, this volume) and if a nuclear pitch accent is not predictable, "n" can be added to mark a nuclear pitch accent (e.g. L+H*n in Italian; Grice, D'Imperio, Savino, & Avesani 2005). If a tonal target is a boundary tone marking the edge of a prosodic unit, a diacritic is added to represent the prosodic unit with which the tonal target is associated. For example, "%" is typically added to an Intonational Phrase (IP) boundary tone (H% for an IP-final tone and %H for an IP-initial tone), "-" to an Intermediate Phrase (ip) boundary tone (e.g. L- for ip-final, -L for ip-initial), and "a" to an Accentual Phrase (AP) boundary tone (e.g. Ha for AP-final, aH for AP-initial). We can use the same convention for the edge tones of a Prosodic Word (PW) by adding a "ω" diacritic as in "Hω" or "ωH."[3] If a tonal target is a phrasal tone, not marking a head or an edge, a tonal symbol without a diacritic can be used (e.g. L, H, LH).

---

[3] A diacritic "w" could be used to refer to a Prosodic Word, but this symbol may be ambiguous as it has already been used among ToBI researchers to refer to a "weak" (undershot) realization of an underlying tone. For example, wL% and %wL are undershot AP boundary tones in Japanese (Venditti 2005) and wL*+H is an undershot pitch accent in Greek (Arvaniti & Baltazani 2005).

Tonal categories and diacritics that can be used in describing and analyzing inton-
ation contours are listed in (7). Here, the tone bearing unit (TBU) is a syllable or a
mora. The list includes categories in tonal levels and complexity, tone-TBU alignments,
and diacritics marking types of head tone, alignment, and prosodic unit. Languages
will differ in which tonal categories or diacritics are needed to describe their intonation
patterns and also which categories are distinctive or allophonic, i.e. vary contextually.

(7)   Tonal categories and diacritics

    (a) By tonal levels
        H   Fo peak over a host TBU (syllable or mora)
        L   Fo valley over a host TBU
        !H   medium Fo level over a host TBU due to reduced pitch range
        M   medium Fo level over a host TBU, contrasting with higher or lower Fo
        ^H   much higher Fo than the default H level (top-line of declination)

    (b) By tonal complexity
        LH   Fo rising over a host TBU
        HL   Fo falling over a host TBU
        LHL   Fo rise-fall over a host TBU
        HLH   Fo fall-rise over a host TBU
        LHLH   Fo rise-fall-rise over a host TBU
        HLHL   Fo fall-rise-fall over a host TBU
        LM   Fo rise to mid level over a host TBU[4]

    (c) Diacritics representing tone–TBU alignment and timing
        *   a starred tone is aligned with a head TBU (syllable or mora),
                e.g. L+H* vs. L*+H
        <   delayed peak[5]
                e.g. H*< (e.g., Bininj Gun-wok ToBI, Bishop & Fletcher 2005; in
                English ToBI, "<" is labeled over an actual Fo peak after the host
                syllable to mark a phonetically delayed peak, e.g. H <.)
        >   early peak, e.g. >H
        +   boundary between two TBUs[6]
                e.g., L+H (rising Fo over two TBUs); H+L (falling Fo over two TBUs)

---

[4] Two types of falling tones, ML (Fo fall from mid to low) or HM (Fo fall from high to mid), can be used
if any of these contrasts with HL.
[5] In Spanish-ToBI, the label L+>H* represents a rising pitch movement with a delayed Fo peak after an
accented syllable. Here, the symbol ">" *before* H* may be confusing to some, as it is the same as the diacritic
meaning "early peak" in the original English ToBI as well as various later ToBIs of other languages.
[6] A boundary symbol "+" can also separate targets in a tritonal sequence (e.g. LH+L for Fo rise over one
syllable and fall to the next syllable; L+HL for Fo rise over two syllables and fall to the second syllable) if the
timing of the fall or rise is distinctive.

Comp. by: PG2846   Stage : Revises2   ChapterID: 0002012552   Date:20/11/13   Time:13:46:38
Filepath:d:/womat-filecopy/0002012552.3D518

OUP UNCORRECTED PROOF – REVISES, 20/11/2013, SPi

:   prolonged tone target
e.g. H:% (prolonged High tone boundary)

(d) Diacritics representing tone types and prosodic units
(T stands for a tone, either a single tone or a complex tone)

T*   the tone T is associate with a head TBU, stressed or pitch accented,
e.g. H*, L*, LH*, L+H*, L*+H, HL*, H*+L

Tf   the tone T marks focus, different from neutral T

Tn   the tone T is Nuclear Pitch Accent (when not predictable) or is a
phrasal tone marking the word with nuclear accent
e.g. L+H*n, Hn

Ta   the tone T is an Accentual Phrase-final boundary tone,
e.g. Ha, La, LHa, L+Ha

aT   the tone T is an Accentual Phrase-initial boundary tone,
e.g. aH, aL, aLH

T-   the tone T is an Intermediate Phrase-final boundary tone,
e.g. H-, L-, LH-, HL-

-T   the tone T is an Intermediate Phrase-initial boundary tone,
e.g. -H, -L, -LH, -HL

T%   the tone T is an Intonational Phrase-final boundary tone,
e.g. H%, L%, LH%, HL%, LHL%, HLH%, LHLH%

%T   the tone T is an Intonational Phrase-initial boundary tone,
e.g. %H, %L

The labels in (7) could also be used as "temporary" labels as a guideline for deciding tonal categories and symbols when analyzing F0 contours in the AM framework before finalizing distinctive categories of the target language. By examining more data, various tonal categories could turn out to be contextual variants of an underlying tone; thus, by labeling all the variants, we may find the contexts that trigger the variant forms. In fact, this is the reason why some ToBI systems include labels that are not distinctive (e.g. downstep pitch accent in Greek ToBI, Arvaniti & Baltazani 2005; ">" in American English ToBI; the phonetic tone tier in Korean ToBI, Jun 2000, 2005a). ToBI is a transcription system of prosody at a phonological level, but it is also a tool for testing and evaluating hypotheses to improve the intonation model, as well as a tool for observing the signal and for creating a communal corpus (Beckman et al. 2005).

If the intonation of various languages is described in the same framework using the same terminology, symbols, and conventions, we could compare prosody across languages more accurately and easily, developing a better model of prosodic typology. A model of intonational phonology should also provide descriptions of the phonetic realizations of each tonal category together with the contexts where the

surface form is realized. This will help us to fully understand the tonal category and avoid the problems of proposing a typology based on the comparison of the abstract categories only (see Ladd 2008).

16.5 Conclusion

In this chapter, we have described methods of studying intonation in the framework of an AM model of intonational phonology, from data collection to identifying tonal categories and finding a prosodic structure. Data collection procedures are described separately when we collect data in the lab and in the field. The suggestions on data design provided in the chapter are mostly based on our own experience and should only be taken as a rough guideline or basic beginning step to start studying intonation. This is so because intonation and prosody can differ widely across languages and even across dialects of the same language depending on the language's or dialect's unique morphosyntactic and semantic/pragmatic information as well as phonological properties. For example, a language may not have a word longer than four syllables, may have complex lexical phonology, or may have a limit on word order or order-specific prosody. However, unlike the data designing procedure, the criteria to decide tonal categories and prosodic structure should apply to all languages if analyzed in the framework of intonational phonology. We hope the methodology and criteria proposed in this chapter can guide students and researchers working on the intonation of any language within the framework of intonational phonology.