

Tone inventories and tune-text alignments

Mary Beckman (*The Ohio State University*)

Annual meeting, Society for Pidgin and Creole Linguistics, Albuquerque, 6-7 January 2006.

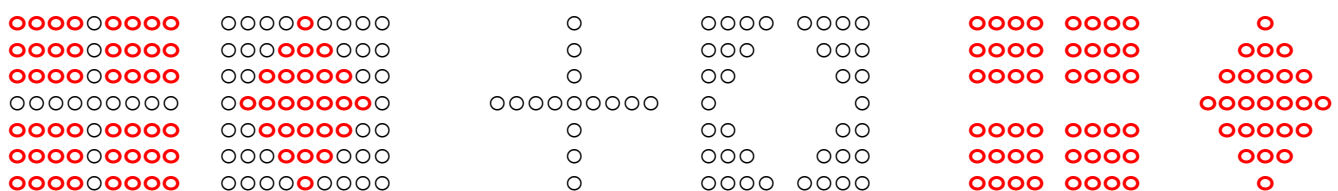
Abstract — In Autosegmental-Metrical accounts, the dichotomies that underlie most traditional typologies of tone and accent can be restated as two questions concerning tone inventory and tune-text alignment. First, are the pitch contours that contrast short utterances composed primarily of patterns specified in the lexicon (so-called “tone languages”) or are they morphemes in their own right (“intonation languages”)? Second, what determines tune-text alignment at the lowest levels of the prosodic hierarchy; are some tones anchored to rhythmically prominent syllables within focused constituents (“stress languages”), or are all tones anchored to prosodic group edges or to rhythmically undifferentiated culminative syllables (“pitch accent languages”)?

0. Preface — “The prosodic systems of several Creole languages have now been characterised as ‘hybrid’ systems since they combine elements of different types of prosodic systems. This is in stark contrast to the traditional typologies of tone, stress-accent or pitch-accent. For example, Saramaccan has been shown to have both a lexical tone system and a pitch-accent system; Papiamentu is argued to have both a pitch-accent system and a stress system. Given what we know about the grammatical systems of these languages, it is not unreasonable to suggest that these prosodic systems are also the result of contact-induced change rather than internal development. At the same time, however, the implications for Creole formation have not been fully explored. To complicate matters, on the other end of the spectrum are Creole languages like Jamaican and Sranan which are said to have purely stress systems and Guyanese (and Haitian?) which has a pitch accent system. The differences in the prosodic systems of these languages present interesting problems for discussions on Creole genesis since many of them are historically related and have similarly emerged out of contact between West African tonal systems and European stress-accent systems. The discussion also present challenges to the traditional typology of prosodic systems.” (Gooden, 2005. [Highlighting added.]

1. Background — A unified “Autosegmental-Metrical” account (term from Ladd, 1996).

- **Metrical**: Refers to the intersecting rhythms or layered prosodic constituents projected away from the content features that are licensed at different positions. For example, in all spoken languages, a low-level prosodic constituent syllable can be defined by the alternation between more sonorant features licensed to at the head and less sonorant features licensed to occur only at the edges. In English, a syllable with a least sonorant head (e.g., [ə] or [l]) must group with a more sonorant syllable in a higher constituent (e.g., foot or prosodic word).
- **Autosegmental**: Refers to the specification of content properties that are autonomously segmented — i.e., that project as strings specified on independent tiers rather than being bundled together in association to the metrical positions that license them. For example, in English, changing specifications of place and/or manner can define as many as eight autosegments per syllable, but alternations of [voice] can define only three.
- **Association**: The relationship between an autosegment and the metrical position that licenses it.

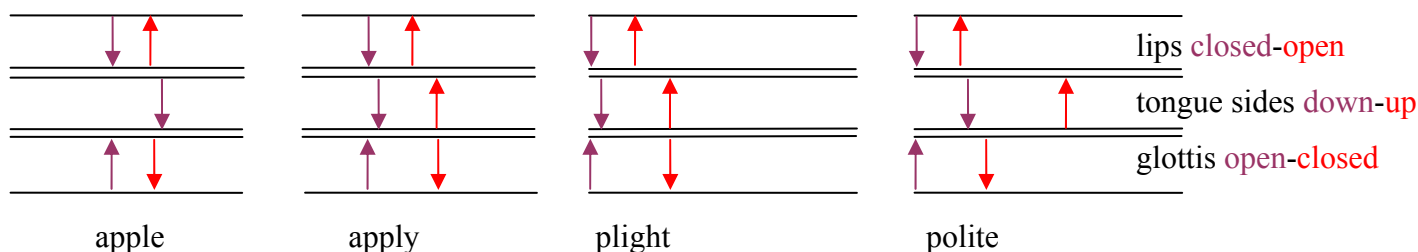
(1) Three different “grammars” for associating [red] vs [black] color “autosegments” to head vs edge positions in the “metrical structures” for [squares] vs [diamond]. Note the “phonological ambiguity” of [squares] if the black edge autosegments are misparsed as the head of a [plus sign] (analogy adapted from Beckman, 1996).



☞ **Challenge 1** — Such parsing ambiguities are especially likely in situations of language contact, where speakers of languages with similar inventories of autosegments but different association grammars can parse the content features completely correctly but misinterpret (or reinterpret) the metrical structures that they cue. (Typologists should be especially cautious, since field linguistics is typically a situation of language contact. Cf. Hyman, in press: “The question that is relevant in this context is whether we want to typologize according to the properties of prosodic systems or according to the analyses given to them by diverse linguists.”).

- **Alignment:** The temporal relationship between dynamics of autosegments from different tiers that are (partially) co-produced because of association to the same metrical position. Alignment patterns can differ for the same set of autosegments in different metrical structures, and these differences can cue the metrical structures in contrast. For example, the glottal opening and closing gestures for the [-voice] autosegment align with the oral gestures for the sequence of place and manner features for the labial obstruent and lateral approximant in different ways in the English words *apply* and *plight* (where the /p/ marks the beginning of a foot or a prosodic word) vs *apple* (where the /p/ marks a foot-medial syllable boundary).

(2) Alignment patterns for labial closure/release, lateral opening/seal, and glottal abduction/adduction gestures for [pl] sequences associated to four different metrical structures. Note the parsing ambiguity of the configuration in *polite*, where some speakers re-interpret the temporal delay between the “lips open” gesture of the [p] and the “tongue sides down” gesture of the [l] as an epenthetic reduced vowel.



- **Sonority scaling:** The autosegments that mark syllable structure are ordered on a continuum from the most open vowels (which are the loudest periodic autosegments) to the most obstruent-like consonants (typically stops with silent closures followed by sharp, transient release bursts). The properties that contrast the category types along this continuum from most to least sonorant segments can be “scaled up” or “scaled down” to help cue metrical structures above the syllable. For example, in many dialects of English, /p, t, k/ have longer closures and a pronounced interval of aspiration when they associate to the beginning edge of a foot or of a prosodic word. The aspiration can be lengthened and intensified in words such *plight* in (2) where the [p] associates to the edge of a word that is also the edge of a foot, and so is “stronger” than the [p] in *apply*, which begins a foot but is not word-initial (see Docherty, 1992, inter alia). Also, autosegments that can associate both with edge and with head positions in syllables, such as the [l] in *plight* vs *polite*, typically are longer and more clearly articulated when associated to head position (see, inter alia, Price, 1980; Browman & Goldstein, 1988).

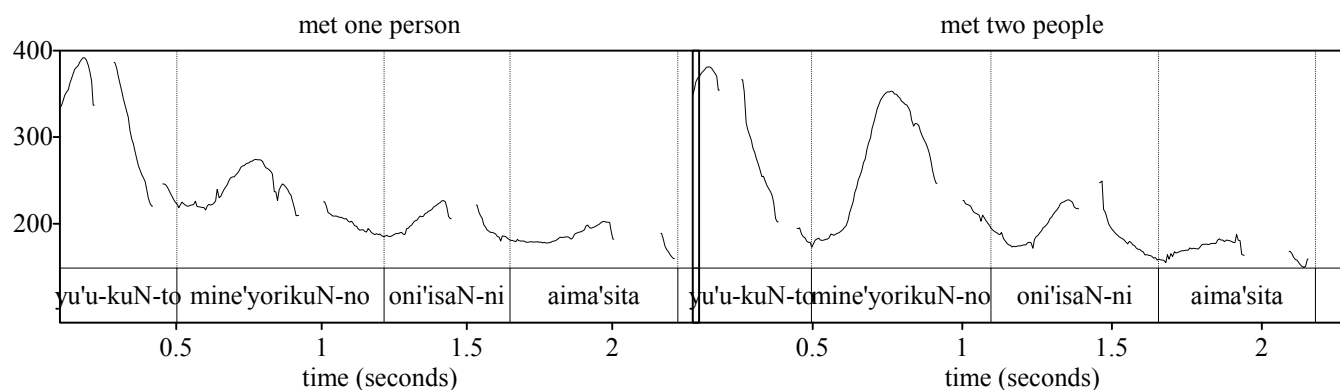
2. Tone in an Autosegmental-Metrical account.

- **Tones:** Tones are the local specification of pitch features as relatively short (i.e., about syllable-length) autosegments independent of the vowel and consonant autosegments with which they are aligned. Tone specifications are inherently arranged in a scale from low pitch (L) to high pitch (H).
- **Tone association points:** Tones differ from other types of autosegments in typically associating to a wide range of metrical positions across and within languages. Like vowel features, tones can associate to syllables or moras (i.e., head positions in syllables). They can also associate to the edges of higher-level constituents such as the accentual phrase of Korean or French and the intonational phrase of English or Cantonese. A tone

or sequence of tones marking the edge of a higher-level constituent in this way is called a boundary tone. In some languages, one particular low-level prosodic constituent or prosodic position is designated as the “minimum tone-bearing unit” (TBU) — i.e., the smallest prosodic constituent that licenses the occurrence of a tonal specification. This term is typically used when the grammar of a language dictates that each constituent at this level must have an associated tonal autosegment.

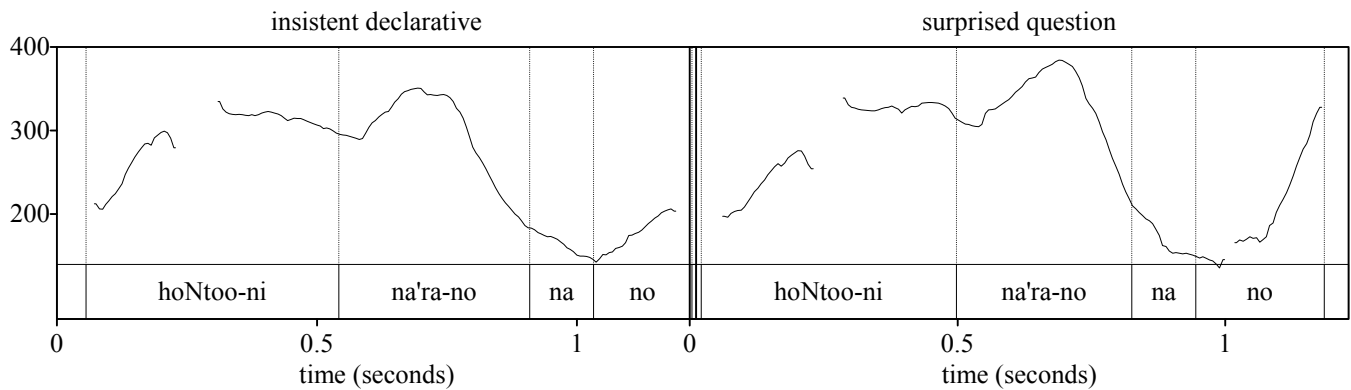
- Pitch range: In addition to such local tone specifications, grammars also specify pitch features for larger constituents, typically referring to parameters of the pitch range — i.e. top and/or bottom values of the tone scale as a whole. For example, many languages mark polarity questions by shifting the entire scale upward over the last intonational phrase in an utterance, so that L tones can be higher than H tones earlier in the utterance. Also, the range can be compressed by shifting the top of the range downward at some fixed point in a prosodic constituent. For example, in standard (Tokyo) Japanese, this downstep occurs at each lexically specified HL, so that each subsequent H tone is lower than the H of the triggering HL sequence until the pitch range is reset to mark the beginning of the next intonational phrase, as in (3).

(3) Fundamental frequency (F0) contours for two utterances of the sentence *Yu'u-kun to Mine'ori-kun no oni'isan-ni aima'sita*, which can be parsed as ‘I met with the older brother of Yuu and Mineyori’ (one person) or ‘I met with Yuu and the older brother of Mineyori.’ (two people), depending on whether the utterance is produced as one intonational phrase (IP), with three occurrences of downstep, or as two intonational phrase, with downstep blocked at the IP boundary after *Yuu-kun-to* (example from Eda, 2004).

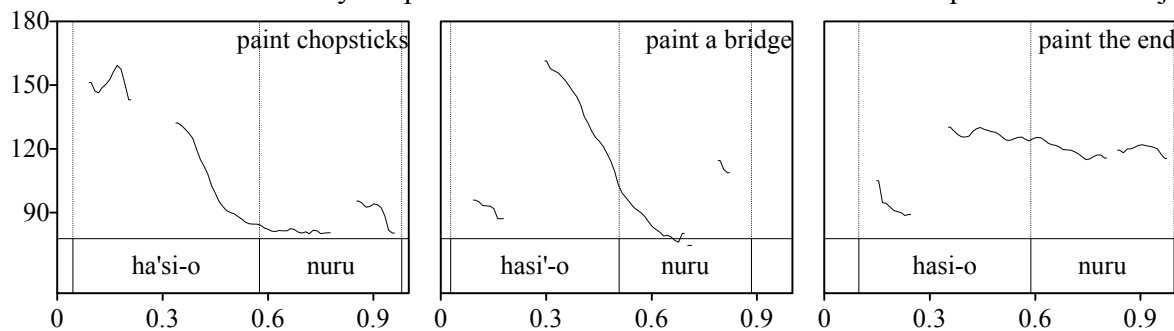


- Morphosyntactic function: Tone specifications can have widely varying morphosyntactic functions both across and within languages. For example, the rise at the beginning of each of the four lexical words in (3) is a fixed LH sequence that is inserted by the phrasal phonology and functions only to mark the beginning of the accentual phrase (AP). The L boundary tone that marks the end of the IP, by contrast, is taken from a rich inventory of pragmatic morphemes that convey such discourse functions as emphatic confirmation of old information vs interrogation, illustrated in (4). The language also assigns a culminative HL tone sequence (an “accent”) to a lexically specified TBU within some APs, and the presence and (if present) the location of this TBU contrasts ordinary lexical words, as illustrated in (5). The first AP (the object NP) in each of the first two sentences in (5) contains an accented lexical item, but the first AP in the last sentence in (5) contains no accented lexical item, hence the lack of downstep on the verb *nuru*. Like Tokyo Japanese, Beijing Mandarin also has IP boundary tones that are pragmatic morphemes in their own right, illustrated in (6), as well as four contrasting tone patterns that associate to syllables in lexical words, illustrated in (7).
- Tone scaling: As with the sonority scale, the relevant acoustic property that defines the tone scale (i.e., F0) can be manipulated to cue higher-level metrical structure and to convey different morphosyntactic functions. For example, in standard (Tokyo) Japanese a L that marks the edge of an IP is lower in the local pitch range than the L that marks the edge of an IP-medial AP. In standard Mandarin Chinese, the L of tone 3 (“dipping tone”) is lower than the L of tone 2 (“rising tone”).

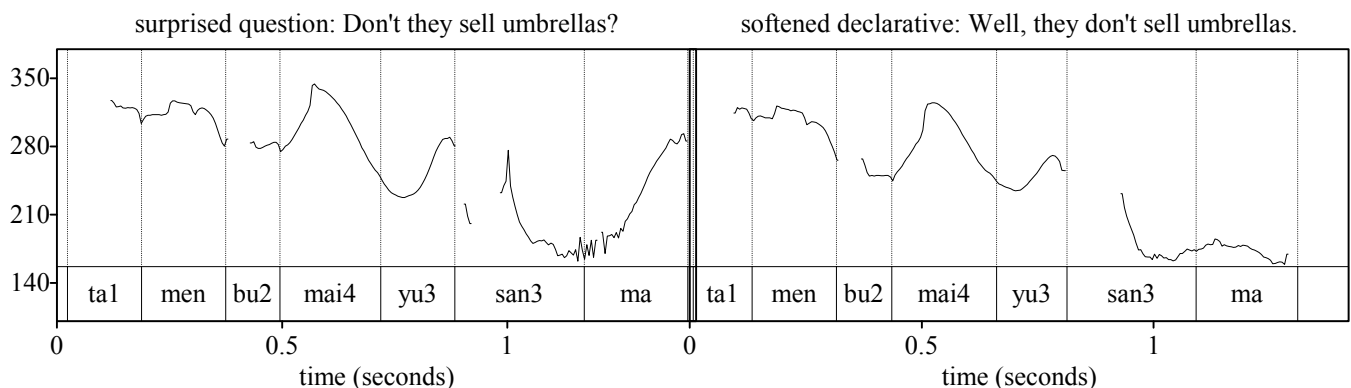
- (4) F0 contours for utterances of the sentence *Hontoo-ni Na'ra-na no.* 'Really the one from Nara.' contrasting two IP boundary tones in Tokyo Japanese. (See Venditti et al., 1998. Example utterances are from Eda, 2004, who shows that L2 speakers and L1 Japanese speakers of other dialects Tokyo have difficulty interpreting these tunes.)



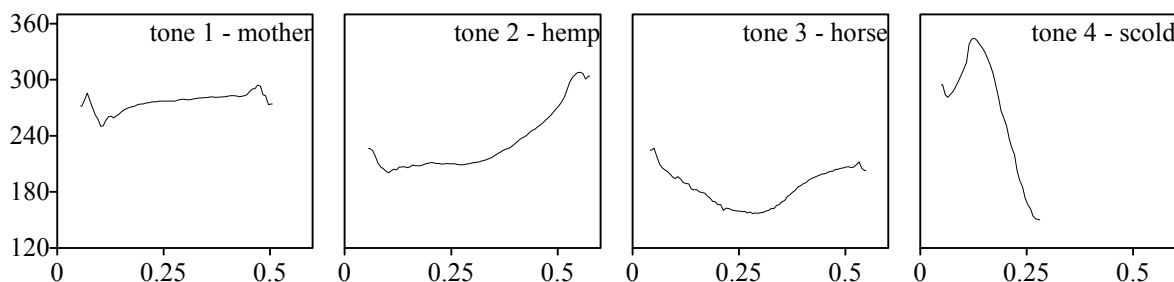
- (5) F0 contours for three Tokyo Japanese sentences that contrast in *akusento* pattern of the object noun AP.



- (6) F0 contours for two utterances of *Ta1-men bu4- mai4 yu3-san3 ma.* 'Don't they sell umbrellas.' produced with contrasting boundary tones by a Beijing Mandarin speaker (from Lee, 2000, cited in Peng et al. 2005).



- (7) F0 contours for citation form utterances of four morphemes that share the same [ma] CV autosegments but differ in tone, produced by a female speaker of (Taiwan Guoyu) standard Mandarin.



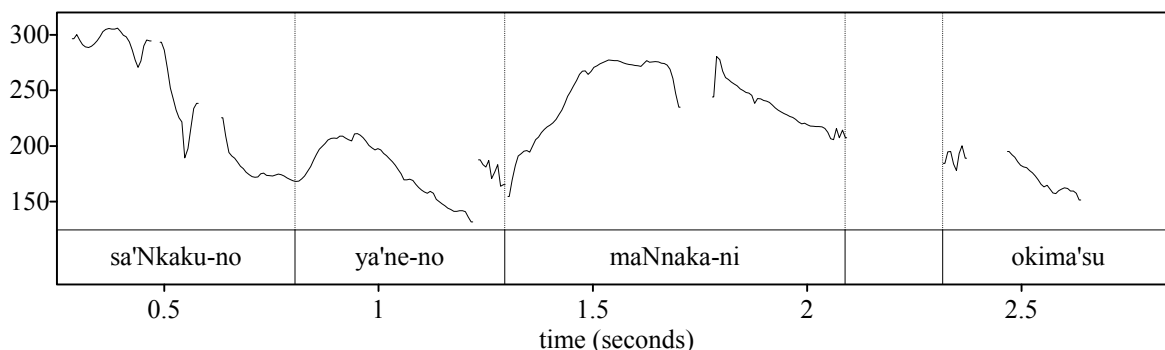
- “Tone language”: In traditional typologies, the tone patterns that contrast in (5) and (7) are singled out as constituting “tone proper.” For example, Hyman (in press) says, “A language with tone is one in which an indication of pitch enters into the lexical realization of at least some morphemes,” and he explicitly excludes the kind of tonal morphemes that contrast in (4) and (6) on the grounds that they are “post-lexically assigned” to the sentence in its discourse context. An “intonation-only language” then would be one in which an indication of pitch enters into the realization of only pragmatic morphemes such as the boundary tones in (4). Hualde et al. (2002) show that some dialects of Basque (a language with a Tokyo type tone system) have become intonation languages of the same type as neighboring dialects of Spanish (n.b. different from French).

☞ **Challenge 2** — Note that intonation languages can use tone to contrast pragmatic functions in different ways (e.g., Gussenhoven, 2004, pp. 45-46), and the ways to be a tone language are even more varied (Hyman, in press, inter alia). For example, Limburgian Dutch, Kyungsan Korean, and Panjabi are tone languages, whereas standard Dutch, Chonnam Korean, and Hindi are intonation languages. The broad cut alone is too coarse-grained to be useful in differentiating contact-induced change from “normal” internal variation.

3. Edge prominence versus stress in an Autosegmental-Metrical account.

- Strong edges: The pitch range reset at the beginning of higher-level constituents such as the IP in Tokyo Japanese often coincides with hyperarticulation of consonants and vowels. That is, consonant and vowel autosegments can be scaled toward the edge values of the sonority scale. The beginning of the IP is associated with more stop-like pronunciations for /b, d, g/ and longer release intervals for /p, t, k/, by comparison to AP-medial /b/ and /d/ (which are very lax and sometimes even approximants) and to AP-medial /g/, which is usually [ŋ].
- Marking focal prominence: These characteristics can be exaggerated even more at the beginnings of constituents that are in focus in the discourse structure (e.g., “new information”). That is, intonational phrasing (and dephrasing of subsequent material) can be used as a focusing mechanism in Japanese (and in many languages). For example, in the utterance in (8), there is a particularly steep rise at the intonational phrase between *ya’ne-no* ‘roof’ and *maNnaka-ni* ‘right smack in the middle’ — which is the new information in this discourse context.

(8) F0 contour for Tokyo Japanese utterance (*Sono pi’Nku-no ma’do-o*) *sa’Nkaku-no ya’ne-no maNnaka-ni okima’su*. ‘I am placing (that pink window) smack in the middle of the triangular roof.’ (Example from the ‘house-building’ task used to elicit spontaneous narrative in Venditti & Swerts, 1996.)



- Pitch accent (definition 1): Note that this exaggerated pitch range and hyperarticulation at the edge of the focal constituent should not be confused with the phenomenon of “accent” in the language. In a Tokyo-type system, “accent” refers to the lexically specified TBU for the H of the HL word melody, which has a culminative distribution in the AP. The words *sa’Nkaku* and *ya’ne* then are “accented” despite being out of focus, whereas *maNnaka* is “unaccented” despite the focal prominence imparted by the steep LH edge rise.

- Relationship to tone in “tone languages”: In some languages, the tones that associate to stressed syllables at some high level of the stress hierarchy are not drawn from an inventory of paradigmatically contrasting pragmatic morphemes, but instead are lexically specified autosegments. In Beijing Mandarin, for example, the tones that associate to stressed syllables at a level analogous to level 3 in (10) are the inventory of four lexical tone patterns that contrast the morphemes in (7). In Stockholm Swedish, similarly, there are two tone shapes that mark stressed syllables at a level homologous to stress level 2 in the English hierarchy in (10), but these pitch accents are lexically specified and contrast words rather than pragmatic morphemes. For example, both *anden* ‘the duck’ and *anden* ‘the ghost’ have primary stress on the first syllable, but the pitch accent that marks this stress on ‘the duck’ is H+L* (i.e., falling onto a L tone target on the stressed syllable) whereas the pitch accent on ‘the ghost’ is H*+L (falling from a H tone target on the stressed syllable). (n.b. In citation utterances, there will be a peak on the second syllable as well, but this is a “phrase accent” that marks the last word in the IP or in a constituent bearing focal prominence — see Bruce, 1977; 1990.)

4. The “obligatory head” constraint.

- “Obligatory head”: In all languages that have been identified to date as having a prominence hierarchy that is structurally equivalent to the one described for English in (10), there is some level of stress that is part of the definition of the prosodic word (see Hyman’s, in press, “obligatory head” constraint). In English, for example, a pitch accent can only associate to a syllable with a full, unreduced vowel and every prosodic word must have (at least) one such syllable. When a function word such as *the*, that normally is pronounced with a reduced vowel (or even with no clear vowel nucleus), is put into focus, it is given a nonce pronunciation with the full vowel that occurs in *heat* or in *hut*. In Swedish, every prosodic word must have a syllable that bears a pitch accent. In Mandarin Chinese, every prosodic word must have a syllable with a fully realized lexical tone. Syllables without lexical tone are either suffixes, such as the *-men* suffix in the word *tal-men* ‘they’ in (6), or function words, such as the negative marker *bu* or the particle *ma* in (6). In fast speech, the vowels in such toneless syllables can be reduced or even deleted.
- Stress language: Depending on the typologist, this term can mean any one of following. (a) It can refer any language that has a hierarchy of syllable prominences structurally like the one in (10). (b) It can refer to the subset of languages in (a) that define the prosodic word in terms of an “obligatory head” syllable that is marked for prominence at some high level of the hierarchy. (c) It can refer just to the subset of languages in (b) in which the tones that associate to stressed syllables come from an inventory of pragmatic morphemes.
- Phonotactics of stress: In many languages, the “obligatory head” is constrained to occur at a fixed position or within a fixed range of positions relative to the prosodic word edge. For example, in Greek, the stressed syllable (the syllable to which pitch accents can associate) must occur on one of the last three syllables of the word. In English, the majority of words have stressed first syllables at the level where stress is marked by an associated full vowel. In Shanghainese compound words, lexical tones on all syllables but the first are deleted (and high vowels in non-initial syllables can be devoiced or deleted).
- ☞ **Challenge 3** — Because languages with stress systems often use a constrained distribution of stress to demarcate lexical words, the misparsing of edge prominence and pitch range reset as markers of stress in languages without stress accent is a very common contact-induced reanalysis. And because many field linguists are first-language speakers of languages with syllable-prominence systems, the same misparsing can influence their phonological analyses. (Cf. Hyman, in press: “Researchers who operate under the assumption that all languages have S[ress]A[ccent] may read stress into the phonetic variations they hear or observe instrumentally.”) For example, linguists who are speakers of English and other Germanic dialects often analyze the LH sequence that aligns to the beginning edge of many accentual phrases in Hexagonal French as a pitch accent placed variably on the first or second syllable of the phrase, although the alignment facts do not support this analysis (Welby, 2003).

- **Focus marking:** In English, focal prominence is marked by the association of an IP edge tone (a “phrase accent”) after the last pitch accent in the focused constituent, which blocks any following pitch accents. The last pitch accent is then the “nuclear accent” and that accented syllable is the head of the IP. In Swedish, focal prominence is marked by the association of a H phrase accent and downstep at each subsequent pitch accent (Bruce, 1977). In Beijing Mandarin, focal prominence is marked by a clear realization of the lexical tones in the focused constituent followed by a drastic compression of the pitch range (Jin, 1996).

☞ **Challenge 4** — As noted, the grammar of focus marking can involve manipulations of phrasing, pitch range, tone scaling, and/or tone association patterns. When words are studied in citation utterances, they are “in focus” and it is impossible to sort out which aspects of the utterance melody are due to lexically specified structure and which to the grammar of focus marking. For example, traditional analyses of Swedish word prosody interpreted the phrase accent as a part of the lexical specification of words bearing the H*+L pitch accent. Until higher-level metrical structure and the grammar of focus-marking are better studied across more (and more diverse) languages, any typology of “word prosody” may be premature.

References

- Beckman, Mary E. (1996). The parsing of prosody. *Language and Cognitive Processes* 11, 17-67.
- , & Jan Edwards (1994). Articulatory evidence for differentiating stress categories. *Laboratory Phonology* 3, 7-33.
- Browman, Catherine P., & Louis Goldstein (1988). Some notes on syllable structure in Articulatory Phonology. *Phonetica* 45, 140-155.
- Bruce, Goesta (1997). *Swedish word accents in sentence perspective*. Lund: Gleerup.
- (1990). Alignment and composition of tonal accents: Comments on Silverman and Pierrehumbert’s paper. *Papers in Laboratory Phonology* 1, 107-114.
- Bolinger, Dwight (1981). Two kinds of vowel, two kinds of rhythm. In Bolinger (ed.) (1986). *Intonation and its parts: The melody of language*. Stanford: Stanford University Press.
- Docherty, Gerard J. (1992). *The timing of voicing in British English obstruents*. Dordrecht: Foris.
- Drayton, Kathy-Ann (2004). Course project for Ling 795T (Practicum on Intonation), Ohio State University.
- Eda, Sanae (2004). *Processing of intonation patterns in Japanese: Implications for Japanese as a foreign language*. Unpublished Ph.D. dissertation, Ohio State University.
- Fry, D. B. (1958). Experiments in the perception of stress. *Language and Speech* 1, 205-123.
- Gooden, Shelome A. (2003). *The phonology and phonetics of Jamaican Creole reduplication*. Unpublished Ph.D. dissertation, Ohio State University.
- (2005). Proposal for session on *Prosodic Descriptions of Creole Languages: Implications for Creole Formation*.
- Gussenhoven, Carlos (2004). *The phonology of tone and intonation*. Cambridge, UK: Cambridge University Press.
- Hualde, José, Gorra Elordieta, Iñaki Gaminde, & Raika Smiljanić (2002). From pitch-accent to stress-accent in Basque. *Laboratory Phonology* 7, 547-584.
- Hyman, Larry M. (in press). Word-prosodic typology. *Phonology*.
- Jin, Shunde (1996). *An acoustic study of sentence stress in Mandarin Chinese*. Unpublished Ph.D. dissertation, Ohio State University.
- Ladd, D. Robert (1996). *Intonational phonology*. Cambridge, UK: Cambridge University Press.
- Lee, Ok Joo (2000). *The pragmatics and intonation of ma-particle questions in Mandarin*. Unpublished M.A. thesis, Ohio State University.
- Peng, Shu-hui, Marjorie K. M. Chan, Chiu-yu Tseng, Tsan Huang, Ok Joo Lee, & Mary E. Beckman (2005). Towards a pan-Mandarin system for prosodic transcription. In Sun-Ah Jun (ed.) *Prosodic typology: The phonology of intonation and phrasing*, pp. 230-270. Oxford: Oxford University Press.
- Price, P.J. (1980). Sonority and syllabicity: Acoustic correlates of perception. *Phonetica* 37, 327-343.
- Vanderslice, Ralph, & Peter Ladefoged (1972). Binary suprasegmental features and transformational word-accentuation rules. *Language* 48, 819-838.
- Venditti, Jennifer, & Marc Swerts (1996). Prosodic cues to discourse structure in Japanese. *Proceedings of the International Conference on Spoken Language Processing (Philadelphia, October 1996)*, vol. 2, pp. 725-728
- Welby, Pauline (in press). French intonational structure: Evidence from tonal alignment. *Journal of Phonetics*.
- Pickering, Lucy, & Caroline Wiltshire (2000). Pitch-accent in Indian-English teaching discourse. *World Englishes* 19, 173-183.