An Autosegmental Typology of Tone: And how Japanese Fits in

John Goldsmith
Massachusetts Institute of Technology

1.0 The purpose of this paper is to consider the tone assignment to nouns in several languages and in two Japanese dialects -- Tokyo and Osaka -- as examples to illustrate some principles central to a universal typology that arises out of a theory of tone that I have been working on.[1]

The theory of tone I call the "autosegmental" theory, for reasons we shall see. It originally arose from work in African languages, and then from work on English. Some of its principles are strikingly illustrated in various Japanese dialects.

One fundamental idea we must adopt is one that is shared by most "suprasegmental" theories: that the pitch melody of a word or phrase constitutes an independent linguistic level. The autosegmental theory, very loosely speaking, is a suprasegmental theory; once we get the slightest bit precise, however, it diverges from theories that claim either (a) that the tone melody exists "above" the standard segmental structure (as a prosody, some might say); or (b) that the tone melody is, at some point during the derivation, mapped <u>onto</u> the standard segmental structure, thus losing its status as an independent level. In contrast to both of these approaches, the autosegmental theory proposes that the tone melody is composed of tonemes which are full-fledged segments in their own right, and that they remain so throughout the derivation. An obvious and immediate consequence of this is that there are two simultaneous segmentalizations of the phonological representation: there is one string of non-tonal (standard) segments, and one (parallel) string of tone segments, or tonemes. The substance of the autosegmental theory consists of a way of relating two parallel strings of segments and of strict constraints on possible variations (i.e., tonological rules) in the relation between the two strings of segments.

An example of this description is (1), which represents a two syllable word with a Low-High melody, realized as Low on the first syllable and rising on the second syllable.

(1)

2.0 We'll return to the details of the theory below. I should like to consider a typology that stems naturally from the theory. This universal typology breaks into two sections: (1) a classification of languages by what factors determine the pitch melody of each word or phrase; and (2) a classification of how the pitch melody, once determined, is then realized on the phrase formally. These two questions form the scaffold of my proposal and this paper. For any language in general,

then, there will be a limited number of ways that the pitch pattern of a word is determined, and a limited number of ways that this melody is realized on the syllabic structure.

2.1 To keep matters fairly simple, I shall restrict the discussion to the tone pattern of nouns. For nouns, we find three possibilities of tone information associated with them in the lexicon:

(1) Languages can specify nothing at all. English is an example of a language like this. There is no information in each lexical entry for the tone of a noun in English.

(2) The second possibility is that there are a limited number of tonal melodies -- not all combinations of underlying tones in a particular language may constitute a possible tone melody for a noun. Rather, in this case of "restricted" melody classes, the lexicon marks each individual noun as to which melody pattern it goes with. Leben's analysis of Mende, or Edmundson and Bendor-Samuel's analysis of Etung show this going on in African tone languages.

(3) The third possibility is where every noun has a tone melody that is simply marked as such in the lexicon. Igbo is such an example. For any noun you pick, the lexical entry for that word specifies completely what the tone melody for the word is -- Low High Low, it might read. This differs from languages like Mende in the second class -- there, the lexical entry for a word would say "Class A", or something like that, and then when you went to a look-up table in the grammar, you'd find that Class A was the melody High-Low.

2.2 Now in this three-way break-down (which is a commonplace these days in one form or another), we have made no use of the distinction between tone languages and accentual languages. That distinction, a useful one for linguists, gets covered in the second section of the typology. If the first distinction may be phrased, "What sort of melody specification is there in the lexicon?", the second may be put, "How does the melody get realised?"

What sort of melody specification is there in the lexicon?	How does the melody get realised?		
		accentual	non-accentual
1) None: English, French, German, Sanskrit, Tokyo Japanese	spread from accent	English Sanskrit Osaka	(logically impossible)
2) Melodic classes: Mende, Norwegian, Etung, Osaka Japanese	spread from word boundary	Tokyo Garda?	Mende
3) Lexical tone specification: Igbo			Igbo

When we consider the second question, we find that there are three independent dimensions along which languages can vary. The first is whether the language is accentual. If a language is accentual, then it distinguishes one of the syllables as perceptually prominent. Linguistically this consists of an abstract mark which

I'll call a "star" (*) on one of the syllables. The assignment of the star to one of the syllables can be either accomplished through rules or through, literally, lexical marking (or, of course, by some combination of the two).

But just sticking some abstract "star" somewhere does not constitute a melody. The second dimension along which languages can vary concerns how they take the tone melody (which was itself determined by the mechanisms we discussed a minute ago) and link it up to the syllable structure of the word. And for this there are two choices, as I noted in the chart above: either you start from the word boundary or the accent mark as you spread out the tone melody over the syllables (or the syllables over the tone melody, if you wish to view it that way). Third, the question arises typologically concerning whether you do this spread leftward or rightward, independently of where it starts from.

2.21 Let's consider the case of spreading from the accent first. English just so happens to be one of the languages that works like this. In the word archipelago, for example, the third syllable -- the antepenult -- gets the star, what is normally called the main stress. The tone melody for English words said in isolation is Mid-High-Low (MHL), if said with neutral intonation. The star, or accent, is on the High. Note this: English is accentual, nobody doubts that -- but that means not only that one syllable is marked as accented, but that one of the tonemes in the tone phrase has the star -- the accent mark -- on it. There is everywhere perfect formal symmetry between the tone level and the syllable structure.

To get the pitch melody of archipelago, then, you hook up the starred tone (the H) to the starred syllable, and then spread the M and the L to the left and right so that every syllable gets a tone. These association lines, as we may call them, indicate which pitch the syllable is uttered at; and these lines cannot cross, or else you'd obviously mess up the melody.

(2)

Osaka Japanese works this way, too. We'll get to those facts below in detail; for the moment, we rest content that the basic formal principle is clear. You hook up the starred syllable with the starred tone, and then hook up tones and syllables to the left and right side of the accent to ensure that all tones and syllables have association lines attached to them.

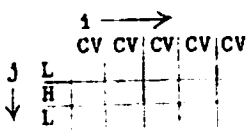
2.22 Now let's consider a more complicated example -- left-to-right spreading from the word-boundary in Mende. These examples come from Will Leben's thesis. The basic idea is again simple, though quite different from English. Mende is not an accentual language, so it has no stars on either its syllables or its tones. Given a tone melody,

you match the tones up with the syllables one to one starting from the left -- the first tone with the first syllable, the second tone with the second syllable, and so on. If there are excess tones (more tones than syllables), then they get associated with the final syllable by a universal well-formedness condition already alluded to, which says that all tones must be hooked to syllables unless that would make these association lines cross each other. The converse holds too, mutatis mutandis, if there are more syllables than tones: the right-most tone associates with all the extra syllables because the well-formedness condition insists that all syllables be associated with at least one tone.[2]

If the idea is clear, let's run through it again more formally. The language-particular rule that needs to be written for the example from English is trivial: it amounts to associating the starred tone with the starred vowel (see (1$) below). All the rest is done by universal convention (i.e., the well-formedness conditions). The case from Mende is a bit less obvious, to be sure. To formalize the "one-to-one" mapping, we shall appropriate the notion of rule schema as set forth in Vergnaud and Halle's work on variables in phonology.[3]

We set up the phonological segments at the top of a sequence of columns, and the tonological segments at the left of a sequence of rows, like (3).

(3)



(for purposes of exposition, I'm going to pretend the consonants aren't even there. The careful reader may correct this for his or herself).

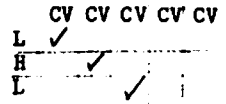
The assignment rule then takes the form of (4).

$$(4)\quad V \approx T \;/\; \# X \begin{matrix} V \\ T \end{matrix} \qquad (\text{"}\approx\text{" means "is associated with"})$$

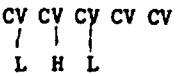
This is to be interpreted in what I take to be the natural way, given the way Vergnaud and Halle's "real" variables work: we produce an algorithm that checks off a certain number of the boxes in the matrix (3). These boxes then represent pairs of vowels and tonemes (column and row), and we apply the structural change of the rule to these pairs of toneme and phoneme, which means in this case to hook them together.

In particular, the algorithm is to check off the box (i,j) (think of the indices as coordinates as indicated in (3)) where $i > i'$, and $j > j'$ for all other checked-off boxes of the form (i',j'). This algorithm is reactivated successively until it cannot do anything anymore.

In the case of (3), this will give us (3').

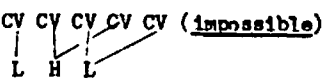
(3') 

Or, written more perspicuously, this yields the output (5).

(5) 

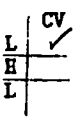
The assignment rule having applied, the well-formedness condition comes into effect (it would have made no sense before this, of course), and this requires that the final two syllables becomes associated with some toneme. But clearly they can become associated with only the final L or else lines will cross, which is strictly forbidden. Thus we end up with (6); situations like (7) could not arise.

(6) 

(7) 

Conversely, given an input like (8), the rule will apply as in (9), giving (10), which becomes (11) by the well-formedness condition.

(8) 

(9) 

(10) 

(11) 

Here is a table of representative Mende words and their melodies (from Leben 1973, p.64).

(12)
H pelé, kó
L bèlè, kpà
HR kénya, mbú
LH nĭka, nàvo, mbâ
LHLnikíli, nyàhâ, mbâ

2.3 To summarize: we have seen how languages like Mende perform tone spreading in a very non-trivial way. Implicit in the formalism we have been using have been just two degrees of freedom: first, we could have spread either from left-to-right (as in Mende), or, just as reasonably, from right to left. Second, the spreading could occur from either of the two tonologically distinguished places in a word: either from a word-boundary (as in the Mende example above) or from the accent. Mende not being an accent language, the second logical

possibility was stifled in the example above, but given an ordering of, first, the rule to associate the starred toneme with the starred vowel, and then, second, the rule to perform the one-to-one associating, we'd end up with mapping <u>from the accent</u>, rather than from the word boundary.

In fact, we find all these possibilities, I believe. Some we shall see below; I shall just briefly mention Sanskrit as an example of a language which maps from right to left starting at the accent, as suggested in the preceding paragraph. In short, Sanskrit has an association rule as in (13) (compare it with (4)):

(13) $V \approx T /$ $\begin{matrix} V & X & \overset{*}{V} \\ | & & | \\ T & & S \end{matrix}$

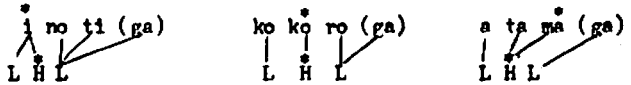
To summarise: in accentless languages, we find one-to-one mapping from the boundary; in accented languages, we have logically and empirically possible cases where the one-to-one mapping begins at the accent. Now the question arises, could there be a language which has mapping beginning at the word-boundary but which was an accentual language? Our typology would suggest that this should be possible -- and it is. It's standard, or Tokyo, Japanese.

3.0 The tone melody in Tokyo Japanese is Low High Low ($\overset{*}{\text{LHL}}$) with the star on the High. To derive the melody of a word with the third syllable accented, for example, you start on the left side. Hook the first tone to the first syllable, making it Low, the second tone to the second syllable -- but now you must keep going, associating this starred High tone with all the intervening syllables until you get to the accented syllable -- for how else would we succeed in associating the accented segments with each other? Then you proceed, assigning the next tone (Low) to the next syllable, and so on, as before.

To put this formally, I'm saying that the tone-association rule in Tokyo is just the rule (4), very common on the African continent. Tokyo, in addition, incorporates the "*" into the array like (3) when applying (4). The "*" dictates a particular modification of the checking-off algorithm: if the algorithm checks off the box of a starred element i, then the checkings must remain in the row or column belonging to i until i has been associated with a starred element. Put in English, this says to perseverate on a starred tone or vowel until it has gotten associated with some other starred element.[4]

We must add a further rule after tone association to derive the correct Tokyo melodies: all contour tones (Rising = $\widehat{\text{LH}}$ and Falling = $\widehat{\text{HL}}$) are simplified to H. It hardly needs to be said that these are among the most common tone rules in creation.

Then we get the following kinds of tone melodies, looking at accented polysyllabics:

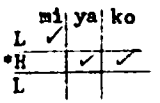
(14) 

For example, a word like kamisori-ga gets its assignment from rule (4) applying as in (15)

(15)

There are accentless nouns in Japanese, too. So far it seems we have made no prediction regarding their tonal behavior; but suppose we simply apply the algorithm as it stands for an accentless word like miyako. First, we set up the matrix (16).

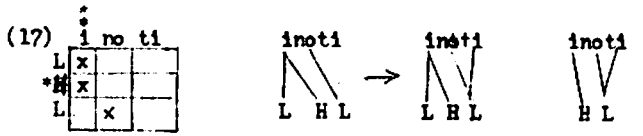
(16)



We check off the box (1,1), which hooks the first L to "mi". Then we check off (2,2), making "ya" associated with H. Continuing with the interpretation of a starred element in this matrix as proposed above, we perseverate with the H until we get to a starred vowel. But we never do, there being none; so the H is associated with all the rest of the syllables.[5]

Before going on to other matters, I should like to point out two immediate consequences of this particular formal approach to tone assignment. First, toneme strings and phoneme strings being formally identical, conditions set up for one must be adhered to by the other. So we would expect situations where the "perseveration" caused by an accent in such an array ran down a column, just as in the cases above it ran across a row. That's just what happens to initial-accented words, as in (17).

(17)

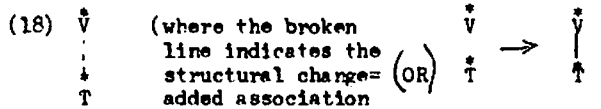


Second, this approach has it as an immediate consequence that when a single melody is associated with a sequence of vowels more than one of which is accented, then the left-most accent "wins". That is, the formalism derives the dominance of the left-most accent from the fact of left-to-right spread.[6]

4.0 Now let's turn to the Osaka dialect. In Osaka, every word must be marked both as to the location of the accent -- just as in Tokyo -- but the noun must also be marked as to which of two melody classes it falls into: either LHL or *HL. Here, as in general, all lexical entries may have at most one accent mark.

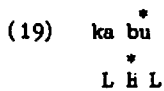
This distinction into two tone classes leads to a bifurcation of the nouns in Osaka which don't have initial accent. This bifurcation

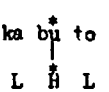
amounts to whether the syllables that precede the accent are all High or all Low. Recall that this is what we'd expect if the associating that is done between the tone melody and the syllable structure is done as in English rather than as in Tokyo. Let's see what that consists of formally, now. This time, you associate the starred toneme and the starred syllable together first, before you do anything else. Such a rule is even simpler than the Mende-type rule (3); this rule is expressed in (18)

(18) 

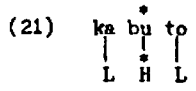
The careful reader will have noticed that a rule like the Sanskrit rule (13) above is fed by (and intrinsically ordered after) rule (18). (18) is also the rule for English.

In the Osaka case, applying rule (18) to a form as in (19) gives us (20).

(19) 

(20) 

By the well-formedness condition, (20) becomes (21).

(21) 

Given a lexical specification as to the melodic class -- whether HL or LHL -- and the location of the accent on the word, rule (18) produces the correct tone patterns if we have the contour simplification rules mentioned for Tokyo -- to wit, a L tone deletes when it is associated with the same syllable as a H tone.

So with initial accent, there are just words like *inoti, which will emerge as in Tokyo with an initial High syllable followed by Low. Contrasting with kabuto, which takes the LHL melody, is *atama, which takes the HL melody, and thus is High on the first two syllables, and Low on the last.

The reality of the melodic class distinction is brought out clearly by the tonal behavior of accentless words in Osaka. When uttered in isolation, our rules make no prediction so far for the tone pattern. This is different from the situation with Tokyo, but for an understandable reason. Tokyo uses the tone-language-like rule (4) to assign tone; Osaka is a true accentual language. Osaka can't apply a tone-melody to an unaccented word for the same reason that the upper right-hand quadrant is impossible in the chart at the beginning of this paper; it doesn't make sense.

Words underlyingly unaccented receive accent on the final syllable.

Thus an unaccented word like susume is realized as " susume " alone and " susume ga" with a particle, because it belongs to the LHL melody class, rather than the HL class. An example of an accentless word in the HL class is miyako, which is realized as "miyako" and "miyako-ga" -- that is, all High.

A striking application of this division into melody classes independent of any marks on the moraic structure arises in connection with the enclitic "gurai". After accented words, gurai loses its accent and participates in the melodic pattern of the stem (thus becoming Low). But when the first word is accentless, gurai doesn't lose its accent, but rather contains the accented syllable of the entire word. However, the melody of the entire phrase is still determined by the lexical specification of the noun in the phrase; words like miyako are High-Low; those like susume are Low-High-Low.

(22) 

5.0 We won't continue with the analysis of Japanese here. The program, however, is to use the notion of typology in a generative grammar more seriously than just to correlate various characteristics. In the case of tone assignment, a language must choose one of the various mutually exclusive possibilities we've looked at. For example, the Tokyo dialect uses rule (4), the Osaka rule (18). The grammars of these dialects then are no more costly for having these rules. The rules of a particular grammar are at least partly necessary just to locate the language typologically -- for example, which type of tone assignment occurs. A language isn't a language without some specification along these lines, so at least as a first approximation we may say that the complexity of a language is not increased by the inclusion of such rules in the grammar. The best situation we could find ourselves in would be if our analysis of the tonal systems of all languages broke down into two parts: one part of a small set of melodic association rules, like (4), (13), and (18), and then supplemental minor rules formalizable within a tightly constrained system, such as the one presupposed here. These minor rules we would hope to constrain to include only (1) contour simplification rules, as described in Japanese here; (2) "flop" rules, which create situations like that in (1) from the corresponding non-contour structure generated by a rule like (4); and (3) deletion and insertion rules subject to the condition that the complexity of the tone melody is not increased.

Such, in any event, is the program; and an optimal theory like that, while not in hand, seems not to be beyond our reach.

* * *

Notes:

1) The notions central to the autosegmental theory have been brewing for quite some time, and I have benefitted from working with Morris Halle in this venture. The particular application to Japanese would

have been impossible without the assistance, both as a linguist and as an informant regarding countless dialects, of Sho Haraguchi. Neither of them agree completely with the proposals made in this paper; the same is true of me. There is some modification of what I'm saying here from what I said orally, most of which is based on further work done by Robert May and myself on Sanskrit. This work was supported by a grant from the National Institute of Health, 5 T01 HD00111-10.

2) In a sense it is too bad that this should be relegated to a footnote, but the topic of the text requires. If you think about the formalism of the autosegmental theory and its well-formedness condition, you will see its very nature makes it essentially suited for expressing -- indeed, explaining -- two kinds of processes: bidirectional "assimilation" or spreading of a feature value over many contiguous segments, up to a segment specified for that feature; and contour-featured segments, like contour-toned vowels, or pre-nasalized stops.

The "spreading" character of autosegments arises from the idea that the autosegments, which sustain the feature ("high tone",e.g.), are segments in their own right, and from the well-formedness condition, which says that all the relevant segments in the other (non-tonal) string must be associated with some tonal segment or other. With the proviso that these association lines may not cross, this accounts automatically for spreading.

The adequate description of contour-featured segments comes from the fact that the contradictory specifications for tone on a vowel are not in fact features of that vowel, but rather two different tone segments associated with the one vowel; and thus both must be uttered while the vowel is uttered. This is what we hear as a contour-toned vowel. A more accurate way of describing this (says the theory) is that during the real-time utterance of the phrase, the switch-over from the tone-segment A to the tone-segment B occurs will we are still inside the non-tone-segment Q. However, it is claimed, none of this constitutes a segment with contradictory features, which is logically impossible because it is incoherent.

The spreading of nasalization in Guarani and its consequent contour nasal segments (i.e., mb, nd, etc.) works precisely in this way, too. The reader is urged to consider Alberto Rivas' enlightening analysis of Guarani in this volume along these lines.

3) In fact, I am trying to work within a framework using only the two types of variables Vergnaud and Halle propose, to the exclusion of iterative application of rules (except at the phonetic level, though they do not discuss this). In such a situation, our hand is forced; we must find a non-iterative formalism to handle one-to-one tone assignment. In the case of the association rule for Mende, a straightforward iterative rendering of the rule is possible, too.

What I said in the text about the spreading of the melody in English being trivial, given the well-formedness condition, is not strictly true. What must be added is the (true) further condition that, when the well-formedness condition demands a spreading of tones to vowels and there is a potential ambiguity as to whether to

spread a starred or an unstarred toneme, always spread the un-starred toneme. This makes sense: it preserves the function of the star as indicating prominence.

4) This same strategy is used, I believe, by Ganda, a Bantu language, which is where the idea came from in the first place.

5) By the well-formedness condition, the final L will associate with the final syllable, creating temporarily a falling tone there, which generally gets simplified to H. Under some conditions, the final vowel will lengthen, and then the L does not delete, but rather is realised as a part of a falling tone.

6) Since writing this paper, I've seen a recent and extremely interesting paper by James McCawley call "Accent in Japanese" which deals with several of the same things this one does. It has affected the present form only marginally, but only because of time limitations. In any event, he speaks of "the predomination of the first accent in a surface structure", and suggests this phenomenon is akin to the Nuclear Stress Rule in English. If the suggestion in the text here is correct, there is no "rule" per se of accent deletion in Tokyo needed to give the appearance of the dominance of the leftmost accent. Anyway, I've suggested elsewhere what I believe is true: the phenomena treated by the "Nuclear Stress Rule" in English is not stress reduction but simply downstep of successive High tones in English (see "English as a Tone Language, Part II")

References

Edmondson, T. and J.T.Bendor-Samuel (1966) "Tone Patterns of Etung," Journal of African Languages 5, 1-6.
Goldsmith, J. "English as a Tone Language" (1974) Unpublished ditto.
"English as a Tone Language: Part II" (1974) Ditto.
Haraguchi,S."Remarks on Accentuation in Japanese" (1973) Unpublished.
Leben, W. Suprasegmental Phonology (1973) Unpublished MIT thesis.
McCawley, J. "Accent in Japanese" Unpublished paper.
May, Robert "Tone Sandhi in Sanskrit"(1974) Unpublished ditto.