Tonal density and classification of tonal languages: Some African cases

Elena Perekhvalskaya
LLACAN, Paris
Valentin Vydrin
INALCO – LLACAN, Paris

„Traditional” comparison of tonal systems

- By the number of tones.
- By type (level tones vs. contour tones).
- By function (lexical vs. grammatical tones).

Carlos Gussenhoven: The tonal density

- In some languages, (nearly) all the tonal bearing units (TBU) have distinct tones. These are languages of a high tonal density.
- In some others, only some TBU carry distinctive tones. These are languages of a low tonal density.
- This parameter allows a quantitative comparison of tonal systems (including „pitch accent” types); their difference will be expressed numerically.
- However, Gussenhoven has not developed a method to define the tonal density in particular languages.

Tonal Density Index (TDI)

- We introduce a Tonal Density Index (TDI) to show more precisely the importance of tones for coding values in a particular language.
- The TDI is the ratio between the number of tonemes (i.e. meaningful tones, or melodies) and the number of segmental units.
- TDI = number of tones tonemes per 100 syllables (or marae) in a random text.

Key notions necessary for the calculation of the TDI

- toneme
- tonal domain
- marked tone and default tone
- toneless syllables/morae
- basic segmental unit (BSU)

Toneme (melody)

- A toneme is a meaningful tone, i.e. a tone (tonal contour) which is relevant for the contrast of lexical or grammatical meanings.
- Rule-induced (i.e., contextually and positionally conditioned and, therefore, predictable) tones are not considered as tonemes.

Tonal domain

- Tonal domain is a segmental chain on which a toneme is realized.
- E.g., in Bambara, tonal domains of variable length:

|*cɛ̀*| ‘man’

|*cɛ̀kɔrɔ*| ‘old man’

|*cɛ̀kɔrɔba*| ‘adult man’

- A floating tone can also represent a tonal domain (of zero length), cf. in Bambara:

|kálo|`| wúlila| *> káló ꜜwúlílá* ‘The moon rose’.

In omnysyllabic languages tonal domains usually coincide with syllables.

Marked and default tones

- In many languages with two contrastive tone levels, one tone (most often, high) can be regarded as marked, and the other one, as default; this situation can be regarded as follows: only high-toned syllables carry tones, other syllables are toneless. In such languages, only marked tones should be counted.
- In such languages too, the notions of toneme and tonal domain are relevant. Only marked tones can represent a toneme; a tonal domain can extend to more than one syllable or more.
- It would be wrong to think that any 2-level tone system are to be interpreted in the terms of “marked tone vs. Ø”. There are languages with two tones where no tone can be regarded as void.

Toneless syllables/morae

- In languages with privative tonal systems (H:Ø, L:Ø), syllables or morae which are not included into the domain of the active toneme are toneless.
- Languages of other tonal types (i.e. those with 2 tonemes or more) can also have toneless syllables. These are TBU to which no tone is attributed on the underlying level.
- Underlyingly toneless syllables receive rule-induced surface tones.

Basic segmental unit (BSU)

- Gussenhoven and Hyman suggested to count the tonal density per morae.
- Theoretically, both types of TDI make sense: the syllabic TDI and the moraic TDI.
- In syllable-counting languages (where syllable = mora), the numeric values are the same in both cases.
- In mora-counting languages, the values of the syllabic TDI and the moraic TDI will diverge.

Problems concerning the syllabic TDI

- In some languages, syllables are hard to discern; ex., in Gokana (Hyman 2009).
- In many language, there is a problem of interpretation of the sequences CVV, monosyllabic (CV:) or dissyllabic (CV-V).
- Our suggestion: the interpretation depends on the monophonemic or biphonemic interpretation of the long vowels. Therefore, it is language-specific.

Problems concerning the moraic TDI

- A „standard” situation in a moraic language: 1 light syllable = 1 mora, 1 heavy syllable = 2 morae.
- There seems to be no unity on the issue of identification of a mora and its relevance in some languages. See below on the problem of segmentation of syllables into morae in Bambara.
- Sometimes, even in closely related languages, heavy syllables are interpreted differently with respect to the number of morae.
- There are languages where 3 or even 4 degrees of syllabic weight are postulated (e.g., some Pular-Fulfulde dialects).

The choice

- Cross-linguistically, segmentation into syllables seems less problematic than segmentation into morae.
- The syllabic TDI, even if its calculation is not devoid of difficulties, seems more universal and provides a better comparability.
- Therefore, we suggest the syllabic TDI as more universal (although the moraic TDI can be also calculated, if necessary).

What precisely do we count for the ITD?

- What is counted for the ITD is the number of tonemes (= tonal domains), i.e. the underlying tones which are represented on the surface level.
- The underlying tones which are erazed (or neutralized) by the rules and do not surface, are not counted.

A very preliminary hypothesis: Three types of tonal languages

- „omnysyllabic languages”: TDI close to 100% (sometimes even more), each syllable has its distinctive tone.
- „tonemic languages”: TDI > 50%, tonemes and tonal domains can be postulated.
- „languages with privative tonal systems” (H : Ø, L : Ø) : TDI < 50%. At its lower end, this type comes close to a toneless language: if a language has regularly only one marked tone per word, such a language can be regarded as accentual, rather than tonal.
- „pitch accent languages”: TDI about 30%. In such languages, only accentuated syllables carry tones.

Case studies. 1. Eastern Dan

- 5 level tones, 1 to 3 contour tones (tones or combinations of tones?)
- replacive grammatical tones
- additive grammatical tones
- profusion of structures CVV, ClVV, CVŋ, CVVV etc. Syllables or sequences of syllables?

Status of the structures CVV, CVVV… etc.: bi- or trimoraic syllables or featural feet?

- If the “moraic” interpretation is accepted, we have to postulate long and extralong vocalic phonemes and several dozens of (phonemic) diphthongs and triphtongs (*o͡a*, *i͡ɤ*, *i͡ɛɛ*, *etc.),* many of them extremely rare (1, 2, 3 occurrences in the vocabulary).
- There are some instances of morphemic boundaries inside the sequences CVV (e.g., *yīʌ́* ‘otter’ < *ýí* ‘water’ + *gɯ́* ‘in’).
- Vowels of the sequence CVV can be sometimes separated by an infix, e.g.: *fɤ̋ɤ̋fɤ̏ɤ̏* ‘light (not heavy)’ → *fɤ̋kɤ̋fɤ̏kɤ̏* ‘very light’ (an intensive form).
- Therefore, these structures are dissyllabic or trisyllabic feet (CV-V, CV-V-V, etc.).

Tonal mapping

- By default, one tone (toneme) is carried by one syllable (= vowel).
- In trisyllabic feet, 1 or 2 different tones are allowed: *ɓúɬ̀ɬ̀* ‘beard’, *ɓãa̋a̋* 1sg. negative focalized pronoun (and not **ɓūɬ́ɬ̀* or **ɓáāa̋*). Therefore, all 3-syllabic feet can be considered as carrying 2 tones (rather than 3); the 3rd vowel (= syllable) is toneless.
- Replacive grammatical tones always substitute lexical tones of the entire foot. E.g.: *dlãã̀* ‘teach’ → *dlãã́* ‘teach\NEUT’. Such grammatical tones can be considered as 1 toneme. The non-initial syllables of a foot are considered as toneless: |*dlã̀ã*|‘teach\NEUT’
- Additive grammatical tones are associated with the preceding syllable which carries, subsequently, 2 tonemes: *gó*`‘sell/INF’.
- If an additive grammatical tone is identical to the preceding lexical tone, it is eliminated: |kɔ̃ +ˋ| → *kɔ̃* ‘refuse/INF’.

TDI for Eastern Dan

- In a text containing 138 syllables, I have counted 146 tonemes.
- The TDI is 105,8.
- In another (more bookish) text, 98.6.
- The average TDI: about 102.
- Eastern Dan is an omnysyllabic tonal language.

Case study 2: Bambara

Syllable, mora, featural foot:

- A foot realized as CCV can be always pronounced as CVCV in a slow speech (e.g., *flà* ‘two’ = *fìla*), therefore, it is considered disyllabic.
- A contrastive long vowel is allowed only in the non-final position in a foot, e.g. *mí:ri* ‘to think’, *bá:ra* ‘work’. The sequence CV: is considered as one syllable, rather than two (no morphological boundary can cut it into two vowels; there are no diphthongs…).

The problem of mora in Bambara

- Two or even three tones can be docked on one syllable, e.g.: [cɛ̌`] ‘man/ART’, [mùsô] ‘woman/ART’. In this case, the vowel of the syllable is realized as long.
- However, the same vowel is realized as short in a different position in a word, ex.: *cɛ̀farinya`* [cɛ̀fàrìnyâ:] ‘courage’, *mùsokɔrɔba`* [mùsòkɔ̀rɔ̀bâ:] ‘old woman’.
- In Bambara, there is no difference in syllabic weight which would affect the ability of a syllable to bear more than one tone.
- Therefore, mora is irrelevant in Bambara.

Tonal system

- Two tonemes, high and low. Contour tones (rising and falling) on one syllable are allowed, but they are regarded as combinations of level tones.
- A toneme is mapped on a tonal domain. The size of a domain most often equals to a word, however, there are words which contain 2 tonal domains: *(kùnna)(bìla)* ‘to reproach’, *(bá)(mànan)* ‘Bambara’.
- On the other hand, one tonal domain can include elements which can be considered sometimes as separate words, e.g.:

(à bɛ) (ń) (dɔ́n) ‘He knows me.’

- A tonal domain can have zero segmental base (a floating low tone):

(mùso)(`) (kúlela) ‘The woman yelled.’

Is Bambara tonal system privative?

- It can be said that in a dissyllabic or plurisyllabic tonal domain, the tone is lexically attributed to the first syllable, and the other syllables are toneless (in fact, they bear surface tones, but these tones are mapped according to general rules of Bambara), ex.:

súruku [súrúkú] ‘hyena’, *bàlaka* [bàlàkà] ‘to precipitate’

- Since (Creissels & Grégoire 1993), the established opinion is that in Bambara (and other Eastern Manding varieties), the low tone is marked (active), and the high tone is a tone by default: *bàlaka*, but *suruku*.
- However, H tone in Bambara has certain degree of activity. Its activity manifests itself in the phenomenona of „tonal compacity” where both high and low tone extends to the right.

Low tone extension: *sò* ‘horse’ + *fìn* ‘black’ + ` ART → |sòfin|`| → *sòfín`* ‘black horse’.

High tone extension: *só* ‘house’ + *fin* ‘black’ + ` ART → |sófin|`| → *sófín`* ‘black house’.

TDI for Bambara

- In a text containing 100 syllables, I have counted 67 tonemes.
- The TDI is 67.
- Bambara is a tonemic language.

Case study 3: Makonde

- Eastern Bantu, group P.23.
- Based on:

Kraal, Peter. 2005. *A grammar of Makonde (Chiminna, Tanzania)*. Leiden: University of Leiden Ph.D. dissertation.

- a 5-vowel vocalic system without phonological length contrast.
- Vowels are automatically lengthened in the penultimate syllable of a word at the end of a syntactic phrase, but also in some other syntactic contexts.
- A syllable with lengthened vowel is bimoraic (therefore, Makonde is a mora-counting language).
- Elsewhere, vocalic sequences are eliminated at the surface level through vowel elimination or their transformation in glides.

Tonal inventory

- Two surface tones, high (H) and low (L).
- Only high tone is active. The low-toned morae are regarded as toneless (i.e., the default low tones are assigned to the toneless syllables and morae postlexically).
- H tones are ascribed to the words lexically, a word can have one, two or no high tones (in the latter case, it is regarded as toneless). The underlying H tones can be regarded as tonemes.
- Verbal stems have no lexical tonal contrasts, they are undelyingly toneless. The superficial tones of verbal word-forms are completely determined by their inflectional and derivational morphology.

Five tonal types of stems in Chiminna Makonde (the underlying H tones are taken into account)

- (A) S1/SF: a H tone on the first and final morae of the stem,
- (B) S1: a H tone on the first mora of the stem,
- (C) SF: a H tone on the final mora of the stem,
- (D) no H: no H tones on the stem,
- (E) S2: a H tone on the second mora of the stem.

Passage from the underlying to the superficial level

- a toneme can be erazed,
- it can be displaced,
- the domain of a toneme can be extended,
- the moraic and/or syllabic structure of a word-form can be modified.

There are lexical and post-lexical rules.

Lexical rules

- Verbs and nouns of C and D classes can be combined with H-toned pronominal prefixes (the other stem classes can be combined only with toneless prefixes).

In the TDI count, these prefixal H tones are regarded as tonemes.

- Meussen's rule (a variant of the Obligatory Contour Principle): When two primary H tones (i.e. tonemic H tones) appear next to each other, the second H is deleted.

In the TDI count, the results of application of Meussen's rule are taken into account.

tu-chí-tóngolá → *tu-chí-tongolá* 'we talked'

Post-lexical processes:
1. Penultimate lengthening (PUL)

- The penultimate syllable of a word-form is lengthened and becomes bimoraic.
- If the primary syllable has a H tone, only the first mora keeps the H tone, and the second mora is toneless (although it can acquire a H tone too when subsequent rules are applied).
- The PUL by itself has no direct impact on the count of the syllabic TDI (it creates an extra mora, not a syllable).

Post-lexical processes
2. Reduction of the complex final syllable and retraction of the final H tone

- The retraction rule, the underlying word-final H tone is automatically shifted to the preceding mora. If the final syllable is light, its H tone is shifted to the second mora of the (lenthened) penultimate syllable.

li-pélepeendé > li-pélepeénde ‘grass (sp.)’

This rule, by itself, has no incidence on the TDI.

- The verb-final syllable can be complex, especially if it is a passive or causative verb, e.g. |-limua| ‘be cultivated’. In the forms with complex final syllables, first, the final H tone is shifted to the preceding more, i.e. the first mora of the word-final syllable; then this syllable loses its initial mora through transformation of the vowel into glide, and the H tone is lost: |ku-límiá| → *ku-líimía* → *ku-líimya* → *kù-lîimyà* ‘to make cultivate’.

This rule diminishes the number of tonemes and therefore influences the TDI.

Post-lexical processes
3. H tone bridge (TB)

- In the stems of the type A. All the syllables between the word-initial H tone and the word-penultimate H tone become H.

|li-pélepeendé| > li-pélepeénde > li-pélépéénde ‘grass (sp.)’

- This rule is not applied to those stems to which the Meussen’s rule has been already applied:

|ku-lí-tóngolá| > ku-lí-tongolá > ku-lí-tongoóla > no TB

- It is not applied to the A-class verbs with a complex final syllable, as far as their final H tone is erased by the preceding rule.
- The Tone Bridge rule results in extension of the domain of a toneme; it does not create new tonemes nor erases them, and it has no incidence on the TDI count.

Post-lexical processes
4. Prefix-H tone shift (PS)

- The H tone of the pronominal prefix is shifted to the stem-initial syllable:

tú-na-pilikaane > tu-na-pílikaane

- If the stem is disyllabic and the penultimate syllable carries a rising tone (LH), the shift does not occur or stops on the tense prefix; this constraints prevents a violation of the OCP (otherwise, two H tones would be adjacent).

vá-na-loóla > vá-ná-loóla

- This rule has no incidence on the TDI count.

Post-lexical processes
Further rules

- Coalescence, resyllabification and OC-H tone retraction (5 different rules)
- Tone doubling,
- Final H deletion,
- Structure Simplification,
- Default L tone isertion.
- Tone assimilation.

The syllabic TDI for the Chiminna dialect of Makonde

- The total number of syllables in the sample text is 125, and the number of tonemes is 41.
- The TDI of Makonde equals 32,8.

Some other languages, by types

- Type 1 (omnysyllabic): Akebu (Southern Guang < Kwa, data from Nadezhda Makeeva), TDI close to 100

Eton (North-Western Bantu, data from Mark van de Velde), TDI 119.

- Type 2 (tonemic): Tibetan (Sino-Tibetan, data from Elena Perekhvalskaya), TDI 52 or 58 (depending on the status of the downstep).
- Type 3 (privative): Navaho, TDI 35.8.
- Type 4 (pitch-accent): Slovene (Slavic) 29, Chokavian (Croat) 26, Aukštaitian Lituanian (Baltic) 25.