The functional load of tone (FLT): High vs. low FLT languages

Nicholas Rolle

Leibniz-ZAS – Berlin, Germany

Theory of Tone (ThoT) workshop
December 1, 2023
A classic observation: Africa vs. Asia

• Classic observation
  ◦ **Asian** languages use tone primarily to distinguish **lexical** items
  ◦ In contrast, **African** languages use tone to distinguish **grammatical** meanings to a much greater extent

• **Rich Grammatical Tone (GT)**
  ◦ Bantu, Mande, Dogon, Gur, Ijoid, Chadic, Nilotic, Central Sudanic, Khoe-Kwadi, Trans-New Guinea, Japonic, Sino-Tibetan, Tanoan, Otomanguean, many isolates (Urarina, Bora, etc.)

• **Poor Grammatical Tone**
  ◦ Sinitic (e.g. Mandarin, Cantonese), Tai-Kadai (e.g. Thai), Austroasiatic (e.g. Vietnamese), and Hmong-Mien

Can this demarcation be refined?

• Can this demarcation be tied to quantifiable differences?
• One possibility involves the **Functional Load of Tone (FLT)**
• For English: “the difference between [d] and [t] is used to distinguish between many different lexical items, so it has a high functional load; there are, on the other hand, very few lexical items that hinge on the distinction between [ð] and [θ], so its functional load is much lower”

Quote: Hall et al. 2016
The functional load of tone (FLT)

• Hall et al.’s 2016 definition of FL applied to FLT (see website [Phonological Corpus Tools](#))
  ◦ A measure of the ‘work’ that tonal contrast does in a language, as compared to other contrasts (e.g. vowel quality, vowel quantity, consonant voicing, etc.)

• FLT – Standard Mandarin (Sinitic: China)
  ◦ mā ‘mother’ | má ‘hemp’ | mǎ ‘horse’ | mà ‘scold’

• Question: Are there high and low FLT languages?

“...different degrees of functional load in the overall phonemic system goes back to the Prague School” – Konnerth 2014:76, citing Mathesius 1929, Jakobson 1931, *inter alia*
Informal statements – High FLT

• “Tone has a high functional load in Supyire. Boys cowherding in the bush often communicate with each other by whistling.”

• “Seenku vocabulary is largely monosyllabic with almost exclusively open syllables, resulting in a high functional load for tone.”

• “Tone carries a high lexical functional load in Ganza.”

• “Unlike some African tone languages, tone is not affected by consonants, tone is stable—it does not shift from one syllable to another, and tone does not down-step or down-drift. The functional load of tone [in Gaahmg] is very high, both in the distinction of words and in the expression of grammatical functions.”

Informal statements – Low FLT

• “The most intriguing aspect of Karbi phonology is the tone system, which carries a low functional load.”

• “Tone functions both lexically and grammatically in [Makary Kotoko] though the functional load of tone is not heavy.”

• “Tone is contrastive both grammatically and lexically in Kifuliiru. Lexical roots which differ only in tone can be found in both nouns (roughly 1%) and in verbs (between 1 and 2%).”

• “pitch accents are contrastive [in Arapaho], and other minimal pairs occur (tecénoo ‘door’ vs. técenoo ‘roll it out!’…). However, such pairs are comparatively rare, and it is extremely difficult to find minimal pairs of nouns based on pitch accent, so the pitch accent system is certainly not equivalent to tonal systems in languages such as Mandarin Chinese.”

The FLT hypothesis

• The FLT Hypothesis:
  ◦ The amount of grammatical tone in a language is negatively correlated with the (lexical) functional load of tone

• Expectation
  ◦ Higher FLT, less GT (i.e. Mandarin)
  ◦ Lower FLT, more GT (e.g. in many African languages)

• Communication-based bias preserves and enhances grammatical marking along ‘under-utilized’ phonological dimensions when lexical contrast would not be significantly compromised
  ◦ I.e. pitch used as inflectional/grammatical/intonational tone

See also Pike’s early statements on Huave tone [huv] (Pike & Warkentin 1961)
Quantifying Functional Load

• Hall et al. (2016): Two ways of calculating functional load

• 1) Change of entropy
  ◦ “Entropy is an Information-Theoretic measure of the amount of uncertainty in a system” (Shanon 1949)
  ◦ What is the change in entropy in a system upon merger of a segment pair or set of segment pairs?

• 2) Minimal pair count
  ◦ “the other is based on simply counting up the number of minimal pairs (differing in only the target segment pair or pairs) that occur in the corpus”
Change in Entropy

• Entropy: \( H = -\sum_{i \in N} p_i \times \log_2(p_i) \)

• We get a baseline of the entropy by first running this measurement for the corpus as a whole
  \( H_1 \)

• We then merge a category and see how much this changes the baseline entropy
  \( H_1 - H_2 \)

• \( p \) is the probability of a word
  ◦ Token-based: multiple by frequency in a corpus
  ◦ Type-based: multiple by 1 (all equally frequent in a dictionary)
The FLT in Hausa

• **A case study from Hausa** [hau] (Chadic – Nigeria)
• “Although tone does not have a functional load comparable to that of many West African languages like Igbo or Yoruba, it does serve to distinguish a number of lexical items” (Newman 2000:599)
• Digitized Hausa lexicon of 10,768 lexemes (Newman’s 2007 dictionary)
• Input into a R-readable spreadsheet by research assistants to Stephanie Shih (Univeristy of Southern California) and Sharon Inkelas (UC Berkeley)
Hausa lexicon

- After curation, **9164 words** as our baseline starting point

<table>
<thead>
<tr>
<th>id</th>
<th>no merger</th>
<th>H vs. L</th>
<th>b vs. t vs. d vs. d'</th>
<th>i vs. e vs. a</th>
<th>a vs. a:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>total</td>
<td>noTone</td>
<td>noCons</td>
<td>noVowel</td>
<td>noVLength</td>
</tr>
<tr>
<td>1 aajizii_H.L.H</td>
<td>aajizii</td>
<td>aaxixii_H.L.H</td>
<td>aajazaa_H.L.H</td>
<td>ajizi_H.L.H</td>
<td>...</td>
</tr>
<tr>
<td>2 algashii_H.L.H</td>
<td>algashii</td>
<td>axxaxii_H.L.H</td>
<td>algashaa_H.L.H</td>
<td>algashi_H.L.H</td>
<td>...</td>
</tr>
<tr>
<td>3 angajeejee_H.H.H.L</td>
<td>angajeejee</td>
<td>axxaxeexee_H.H.H.L</td>
<td>angajaajaa_H.H.H.L</td>
<td>angajeje_H.H.H.L</td>
<td>...</td>
</tr>
<tr>
<td>4 askakkee_L.H.H</td>
<td>askakkee</td>
<td>axxaxxee_L.H.H</td>
<td>askakkaa_L.H.H</td>
<td>askakke_L.H.H</td>
<td>...</td>
</tr>
<tr>
<td>Merger</td>
<td>Unique</td>
<td>$H$</td>
<td>$\Delta H$</td>
<td>% less compared to tone</td>
<td></td>
</tr>
<tr>
<td>------------------------</td>
<td>--------</td>
<td>--------</td>
<td>------------</td>
<td>-------------------------</td>
<td></td>
</tr>
<tr>
<td>All words</td>
<td>8913</td>
<td>13.12170</td>
<td>0.00016</td>
<td>0%</td>
<td></td>
</tr>
<tr>
<td>noAlvPostA</td>
<td>8912</td>
<td>13.12153</td>
<td>0.000016</td>
<td>0%</td>
<td></td>
</tr>
<tr>
<td>noR</td>
<td>8909</td>
<td>13.12105</td>
<td>0.00065</td>
<td>1%</td>
<td></td>
</tr>
<tr>
<td>noH</td>
<td>8901</td>
<td>13.11975</td>
<td>0.00194</td>
<td>2%</td>
<td></td>
</tr>
<tr>
<td>noFH</td>
<td>8899</td>
<td>13.11943</td>
<td>0.00227</td>
<td>3%</td>
<td></td>
</tr>
<tr>
<td>noFW</td>
<td>8896</td>
<td>13.11894</td>
<td>0.00275</td>
<td>3%</td>
<td></td>
</tr>
<tr>
<td>noVel</td>
<td>8875</td>
<td>13.11553</td>
<td>0.00616</td>
<td>7%</td>
<td></td>
</tr>
<tr>
<td>noGlot</td>
<td>8800</td>
<td>13.10329</td>
<td>0.01841</td>
<td>21%</td>
<td></td>
</tr>
<tr>
<td>nuHMV</td>
<td>8780</td>
<td>13.10001</td>
<td>0.02169</td>
<td>25%</td>
<td></td>
</tr>
<tr>
<td>noVoic</td>
<td>8746</td>
<td>13.09441</td>
<td>0.02729</td>
<td>31%</td>
<td></td>
</tr>
<tr>
<td>noVL</td>
<td>8701</td>
<td>13.08697</td>
<td>0.03473</td>
<td>40%</td>
<td></td>
</tr>
<tr>
<td>notone</td>
<td>8393</td>
<td>13.03497</td>
<td>0.08672</td>
<td>100%</td>
<td></td>
</tr>
<tr>
<td>noV</td>
<td>7770</td>
<td>12.92370</td>
<td>0.19800</td>
<td>228%</td>
<td></td>
</tr>
<tr>
<td>noC</td>
<td>2912</td>
<td>11.50779</td>
<td>1.61390</td>
<td>1861%</td>
<td></td>
</tr>
<tr>
<td>Nouns (n=5781)</td>
<td>Merger</td>
<td>Unique</td>
<td>$H$</td>
<td>$\Delta H$</td>
<td>% less compared to tone</td>
</tr>
<tr>
<td>----------------</td>
<td>--------</td>
<td>--------</td>
<td>-----</td>
<td>-----------</td>
<td>------------------------</td>
</tr>
<tr>
<td>total</td>
<td></td>
<td>5674</td>
<td>12.47015</td>
<td></td>
<td></td>
</tr>
<tr>
<td>noAlvPostA</td>
<td>$t\ddagger$ vs. $k^y$</td>
<td>5673</td>
<td>12.46990</td>
<td>0.00025</td>
<td>1%</td>
</tr>
<tr>
<td>noR</td>
<td>$r$ vs. $ɾ$</td>
<td>5671</td>
<td>12.46939</td>
<td>0.00076</td>
<td>2%</td>
</tr>
<tr>
<td>noH</td>
<td>$h$ vs. $\emptyset$</td>
<td>5670</td>
<td>12.46913</td>
<td>0.00102</td>
<td>3%</td>
</tr>
<tr>
<td>noFH</td>
<td>$f$ vs. $h$</td>
<td>5668</td>
<td>12.46862</td>
<td>0.00153</td>
<td>5%</td>
</tr>
<tr>
<td>noFW</td>
<td>$f$ vs. $w$</td>
<td>5665</td>
<td>12.46786</td>
<td>0.00229</td>
<td>7%</td>
</tr>
<tr>
<td>noVel</td>
<td>$k$ vs. $k^y$ vs. $k^w$</td>
<td>5658</td>
<td>12.46608</td>
<td>0.00407</td>
<td>13%</td>
</tr>
<tr>
<td>nuHMV</td>
<td>$i/u$ vs. $e/o$</td>
<td>5651</td>
<td>12.46429</td>
<td>0.00586</td>
<td>18%</td>
</tr>
<tr>
<td>noGlot</td>
<td>$b$ vs. $ɓ$, ...</td>
<td>5633</td>
<td>12.45969</td>
<td>0.01046</td>
<td>33%</td>
</tr>
<tr>
<td>noVL</td>
<td>$a$ vs. $a$:</td>
<td>5632</td>
<td>12.45943</td>
<td>0.01072</td>
<td>34%</td>
</tr>
<tr>
<td>noVoic</td>
<td>$s$ vs. $z$, ...</td>
<td>5602</td>
<td>12.45173</td>
<td>0.01842</td>
<td>58%</td>
</tr>
<tr>
<td>notone</td>
<td>$H$ vs. $L$</td>
<td>5550</td>
<td>12.43827</td>
<td>0.03188</td>
<td>100%</td>
</tr>
<tr>
<td>noV</td>
<td>$i$ vs. $e$ vs. $a$ ...</td>
<td>5251</td>
<td>12.35838</td>
<td>0.11177</td>
<td>351%</td>
</tr>
<tr>
<td>noC</td>
<td>$b$ vs. $t$ vs. $d$...</td>
<td>2291</td>
<td>11.16176</td>
<td>1.30839</td>
<td>4104%</td>
</tr>
</tbody>
</table>
Hierarchy of functional load in Hausa

• Hierarchy of functional load from highest to lowest FL:
  • Consonantal contrast
    • Vocalic contrast
    • Tonal contrast
  • Major oppositions
    ◦ (Voicing, Length, Glottalization, Vowel Height, Secondary Articulation such as palatalization)
  • Phoneme contrasts
    ◦ (f vs. h, r vs. ŋ, etc.)
Quick comparison

- Mandarin/Cantonese: FLT equivalent to that of vowels
- English/German/Italian stress: marginal FL

<table>
<thead>
<tr>
<th></th>
<th>Cantonese</th>
<th>Mandarin</th>
<th>English</th>
<th>French</th>
<th>German</th>
<th>Italian</th>
<th>Japanese</th>
<th>Korean</th>
<th>Swahili</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\text{FL}_{\text{Cons}}$</td>
<td>10.64</td>
<td>13.09</td>
<td>20.82</td>
<td>19.41</td>
<td>15.45</td>
<td>11.12</td>
<td>9.39</td>
<td>11.5</td>
<td>20.0</td>
</tr>
<tr>
<td>$\text{FL}_{\text{Vowel}}$</td>
<td>4.55</td>
<td>3.24</td>
<td>6.7</td>
<td>14.83</td>
<td>4.37</td>
<td>7.61</td>
<td>3.76</td>
<td>3.3</td>
<td>4.11</td>
</tr>
<tr>
<td>$\text{FL}_{\text{Tone}}$</td>
<td>4.48</td>
<td>4.13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\text{FL}_{\text{Stress}}$</td>
<td></td>
<td></td>
<td>.005</td>
<td>.01</td>
<td>.24</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Prospects and challenges

• **Issue 1:** Comparing like with like – **Language systems**
  ◦ “Maybe Mandarin FLT is so different because it is very analytic, unlike languages like Hausa which is more synthetic (let alone Bantu...)

• **Issue 2:** Data sources are not equivalent
  ◦ The Chinese FLT studies cited are based on corpora, where frequency of particular lexical items is controlled for

• **Issue 3:** Counter-examples
  ◦ “My language has no GT but also has virtually no FLT”
  ◦ “My language has lots of GT but also has high FLT”

• **Issue 4:** Quantifying the amount of grammatical tone as well as the functional load of grammatical tone

Recently: functional load of grammatical tone in Grimm 2023
Comparing like with like: Ebira vs. Izon

- **Ebira** [igb] (Benue-Congo: Nigeria) – Verb root generally not targeted by grammatical tone
  - Continuous: $mā̀̄̄ve$ 'I am coming'
  - Completed: $māā̄ve$ 'I came'
  - Interrogative: $má́̄̄ve$ 'did/do I come?'
  - Conditional: $mà̋̄̄ve$ 'if I come'
  - Subjunctive continuous: $mā̄̄ve$ '(that) I should be coming'
  - Conditional habitual: $màáá̄ve$ 'if I usually come'
  - Conditional continuous: $màā̄̄ve$ 'if I am coming'
  - Interrogative subjunctive continuous: $máá́̄̄ve$ 'should I be coming?'

Comparing like with like: Ebira vs. Izon

- **Izon** [ijc] (Ijoid: Nigeria) – Verb root frequently subject to grammatical tone changes, often replacive

- Verb roots contrast in tone in isolation:
  - /dì\(^{\text{LH}}\)/ [dìí] 'look at'

- **Overwriting** of verb tones in context (neutralizes lex. tone):
  - tòbòù\(^{\text{LH}}\) 'child' → [tòbòù ðìí] 'look at a child'
  - tòrù\(^{\text{H}}\) 'river' → [tòrù ðìí] 'look at a river'
  - òndù\(^{\text{HL}}\) bí\(^{\text{L}}\) 'the other' → [òndù bí ðìí] 'look at the other'

Izon: Williamson 1965:11,37,109,102; Rolle 2021
FLT (minimal pairs of 1σ-Verbs)

• **Ebira**: “it is easy to observe some lexical contrasts on monosyllabic items especially verbs of CV syllable structure”

• **Izon**: minimal pairs are difficult to find, e.g. bọọ^LH 'agree' vs. bọọ^H 'pass, be too much'

• Let us only compare like with like: **1σ verbal roots**

<table>
<thead>
<tr>
<th>Lang.</th>
<th># 1σ-Verbs</th>
<th>Unique</th>
<th>No-Tone</th>
<th>Decrease</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ebira</td>
<td>n=194</td>
<td>166</td>
<td>102</td>
<td>38.55%</td>
</tr>
<tr>
<td>Izon</td>
<td>n=200</td>
<td>187</td>
<td>169</td>
<td>9.63%</td>
</tr>
</tbody>
</table>

Ebira quote: Adive 1989; Izon data from Williamson & Timitimi 1983, in Kolokuma dialect
Prospects and challenges

- **Issue 1:** Comparing like with like – *Language systems*
  - “Maybe Mandarin FLT is so different because it is very analytic, unlike languages like Hausa which is more synthetic (let alone Bantu...)”

- **Issue 2:** Data sources are not equivalent
  - The Chinese FLT studies cited are based on *corpora*, where *frequency* of particular lexical items is controlled for

- **Issue 3:** Counter-examples
  - “My language has no GT but also has virtually no FLT”
  - “My language has lots of GT but also has high FLT”

- **Issue 4:** Quantifying the amount of grammatical tone as well as the functional load of grammatical tone
  
  Recently: functional load of grammatical tone in Grimm 2023