Finite state prosodic analysis of African corpus resources

Dafydd Gibbon

Fakultät für Linguistik und Literaturwissenschaft
Universität Bielefeld, Germany
gibbon@spectrum.uni-bielefeld.de

Abstract

The issue of efficient language documentation, particularly with regard to minority and endangered languages, has gained in importance in recent years, as witnessed by several major funding programmes and other human language technology initiatives in the field. An application of finite state technologies to the processing of lexical tone variation in annotated corpora of African languages is described. It is shown that finite state transducers can be constructed which not only provide adequate models for contextual variation in lexical tone (including automatic downstep, downdrift, and tonal assimilations), but also that the transducers provide intuitively satisfying explications of prosodic concepts in 'metrical phonology' in terms of oscillations (iterative transitions). The technique has both theoretical value in formalising typological differences in African lexical tone languages and practical value in automatically generating markup enhancements for concordance-based corpus analysis and for fundamental frequency prediction in pitch modelling.

1. Documentation of African languages

In recent years interest in the efficient and detailed electronic documentation of cognitively and culturally interesting languages has been increasing, independently of the usual commercial interest in language documentation. This interest has been partly motivated by a general scientific and humanitarian interest in the cognitive and cultural diversity of the world's languages, partly by a sense of responsibility for efficiently documenting disappearing languages before it is too late, and partly for reasons of language planning for education and local agricultural and industrial development.

The documentation of African languages brings a number of challenges which are unknown in the documentation of commercially interesting standardised languages; most are unwritten (or have competing orthographies developed by well-meaning alphabetisers), and all are typologically very different from European languages at all levels of description.

One of the issues, the mapping of corpus representations of tonal prosody to representations of lexical tone, is discussed in this paper with reference to a number of African languages.[1] First, the question of lexical tone in African languages and the realisation of 'allotones' is outlined. Second, the technique of modelling tonological rules by means of finite state transducers is introduced, based on the two-tone system of Ewe/Gbe, then the tone model is generalised to cope with further kinds of typologically interesting complexity. Third, corpus modelling issues and techniques are discussed with reference to the bidirectional Finite State Transducer (FST) mapping between lexicon and corpus. Finally, some open questions are discussed with respect to future directions of research.

[1] I am indebted to Firmin Ahoua, Julie Carson-Berndsen, Steven Bird, Doris Bleiching, Bruce Connell, Kofi Folikpo, Ulrike Gut, Nils Jahn, Mark Liberman, Soma Ouattara, Zakari Tchagbale, Thorsten Trippel, Shu-Chuan Tseng and Eno-Abasi Urua for intensive discussion over a number of years of various aspects of the work presented here.

2. Lexical tone in African languages

2.1. Lexical tone contrasts

We start with discussion of the relatively straightforward basic tonal system of Ewe/Gbe, a typical West African register tone language (a variety of Gbe, a group of Kwa languages spoken in Ghana, Togo and Benin). Ewe has register (level) tone contrast in isolated words:

/*àkà*/ 'test' — /*àká*/ 'coal'

The realisation of the tones depends on the context (contextual variations of this kind are sometimes referred to as *tone sandhi*):

- Tones in sequence tend to fall (declination).
- But: the second tone in a LH sequence is phonetically lower than it would be in a HH sequence.
- e.g. the H on the second syllable of /*èkú*/ — 'death' is realised as mid, not as H.

The following is an example of a sequence:

/ *tó*	*gbúé*	*èvè*	*dzè*	*èkú*	*èdò* /
father	old	two	contract+∅	death	disease

"Two grandfathers became seriously ill"

2.2. Tonological rules

These tonological facts can easily be formulated with conventional nonmonotonically ordered (default, elsewhere) phonological rules (cf. [1]):

$$H \rightarrow \left\{ \begin{array}{l} \text{!h / L} \_\_\_ \\ \text{h} \end{array} \right\}$$

$$L \rightarrow \left\{ \begin{array}{l} m \ / \ H \ \_\_\_ \\ l \end{array} \right\}$$

The high and low lexical tones H and L are mapped into a $!h$ (downstepped high) and m (mid) tone, respectively. Taking into account the fact that in each case the tones assimilate, the set of tones can be generalised with variables $\mathcal{T}$, t:

$$\mathcal{T} \rightarrow \left\{ \begin{array}{l} \sigma t \ / \ \neg\mathcal{T} \ \_\_\_ \\ t \end{array} \right\}$$

where σt denotes a sandhi function, and $\neg\mathcal{T}$ means complementarity of values.

2.3. The ‘whole picture’

What these rules alone do not show is the whole picture of what is happening in this tone system. The central property of systems of this kind is *tone terracing*, i.e. a regular descent of the tone pattern at well-defined positions. As a preliminary to formalisation, we introduce a systematic extension of traditional terminology for this kind of pattern.

Terracing is a major tonotactic (tone syntax) property of West African register tone languages. In a sequence of lexical tones such as

$$H_1^\frown H_2^\frown L^\frown H_3$$

there are relational constraints on pitch mapping:

$$pitch(H_1) - pitch(H_2) < pitch(H_2) - pitch(H_3)$$

These constraints fall into two main types:

1. Demiterrace (DT) relational constraint:

$$pitch(H_1) \geq pitch(H_2)$$

   Note that ($pitch(H_1) < pitch(H_2)$), ‘upsweep’, is also observed in some West African languages such as Baule.

2. Terrace (T) relational constraint:

$$pitch(H_2) >> pitch(H_3)$$

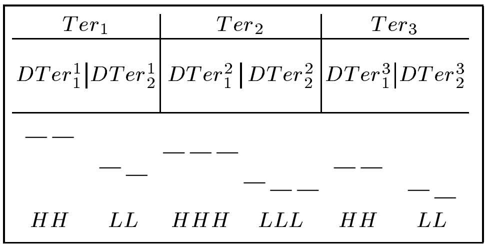

Figure 1: Stylised model of terracing.

The perceptual effect of the terracing constraints is one of tones grouped into sequences of descending ‘terraces’, i.e. H tone sequences followed by L tone sequences, with each H tone sequence at a lower level than the preceding H tone sequence. This is illustrated in idealised form in Figure 1.

2.4. Sequential constraints

The situation is complex, but it appears that we can formulate a number of sequential constraints:

Tone sandhi:

In $H^\frown L$ sequences, L assimilates to m (or even h):

$L \rightarrow m \,/\, H$ _

$L \rightarrow h \,/\, H$ _

Downstep:

IF in a sequence $pitch(H_1^\frown H_2), pitch(H_1) >> pitch(H_2)$
THEN an explanation requires
EITHER a distinct tone (downstepped high), HD
OR a non-surfacing L, L^0, with
the interpretation $pitch(H_1^\frown L^0{}^\frown H_2) = h^\frown !h$.

Independent evidence for L^0 comes from syllable reduction; *downstep* then generalises to H demi-terrace lowering after lexical L tones, whether they surface or not.

Segmental constraint:

There are also constraints on the cooccurrence of tone with nasal and lateral segments in Ewe:

$$L \rightarrow m / \left[\begin{array}{c} \overline{\quad\quad} \\ \left\{ \begin{array}{l} \text{+nasal} \\ \text{+lateral} \end{array} \right\} \end{array} \right]$$

3. Finite state models of lexical tone sandhi

3.1. Applications of finite state phonology to prosody

Since the advent of two-level morphology and finite state phonology over 20 years ago, it has been well-known that phonological rules, under certain assumptions about rule ordering, can be formalised by means of finite state transducers (for an exhaustive overview, see [2]). It has been shown in [3], [4] that this also applies to prosodic rules for intonational prosody, and in [5] that the technique is also valid for lexical tone patterning.

There is additional practical interest in this technique, e.g. in extracting prosodic information from prosodically annotated corpora, as well as more general theoretical interest in language and cognition: prosodic patterning is essentially the distribution of phonetic information over different time domains, and has often been described in *metrical*, i.e. in some sense *rhythmic* terms [6], which are formalised here in terms of cycles in finite state automata.

3.2. Regular expression representation

The first step in providing a model for tonological rules is to translate the rules into regular expressions (specifically: expressions for regular relations rather than regular languages, cf. [7]). The rules given above are modelled by the following regular relation:

$$\{< LH, l!h >, < HH, hh >, < HL, hm >, < LL, ll >\}$$

As with the tone sandhi rules, variables $\mathcal{T},\ t$ can be introduced in order to abbreviate the expressions (the complement operation only affects the h component of the values of t):

$$\{< \mathcal{T}\neg\mathcal{T}, t\sigma\neg t >,\ < \mathcal{T}\mathcal{T},\ tt >\}$$

The following disjunctive regular expression models the tone sandhi rules given above:

$$A|B$$

where

$$A = < H, h > \{< H, h > \{< L, m >< L, l >^* < H, !h >\}^*\}^*$$
$$\{< L, m >< L, l >^*\}$$
$$B = < L, l > \{< L, l > \{< H, !h >< H, h >^* < L, m >\}^*\}^*$$
$$\{< H, !h >< H, h >^*\}$$

The disjunctive regular expression for this kind of cyclic graph is repetitive and inelegant in comparison to the direct statement of the relation. The symbols may be generalised with variables, however:

$$< \mathcal{T}, t > \{< \mathcal{T}, t > \{< \neg\mathcal{T}, \sigma\neg t >< \neg\mathcal{T}, \neg t >^* < \mathcal{T}, \sigma t >\}^*\}^*$$
$$\{< \neg\mathcal{T}, \sigma\neg t >< \neg\mathcal{T}, \neg t >^*\}$$

3.3. Automaton representation of metrical pattern

The regular expression given for the tone patterns is modelled by the finite state transducer shown in Figure 2 in a standard notation.

$$
\begin{array}{l}
< q_0 = q_{start}, \\
\quad Q = \{q_{start}, q_{high}, q_{low}\}, \\
\quad F = \{q_{high}, q_{low}\}, \\
\quad \Sigma = \{H, L\}, \\
\quad \Delta = \{h, !h, m, l\}, \\
\quad f(\delta \times \lambda) = \\
\quad \{ \; tonemap_{ewegbe}(q_{start}, H) \rightarrow h, q_{high}, \\
\quad\quad tonemap_{ewegbe}(q_{start}, L) \rightarrow l, q_{low}, \\
\quad\quad tonemap_{ewegbe}(q_{high}, H) \rightarrow h, q_{high}, \\
\quad\quad tonemap_{ewegbe}(q_{high}, L) \rightarrow m, q_{low}, \\
\quad\quad tonemap_{ewegbe}(q_{low}, H) \rightarrow !h, q_{high}, \\
\quad\quad tonemap_{ewegbe}(q_{low}, L) \rightarrow l, q_{low} \quad \} >
\end{array}
$$

Figure 2: Tone transducer definition.

3.4. Transition network representation

The transition diagramme for the Ewe automaton can be formulated as in case (a) in Figure 3; the additional hc and lc symbols stand for (near-) constant pitch levels.

Although the other notations are formally equivalent, the transition network has the intuitively satisfying property of visualising rhythm directly as an *oscillation*, represented as looping between distinct states. There are evidently two superimposed oscillations: the demi-terrace (local loop on one state), and the terrace (loop between two states).

3.5. Generalisation to typologically more complex cases

The model generalises easily to other typologically related languages. Case (b) in Figure 3 associates the linguistic terminology which is generally in use for the different tonal processes with transitions in the lexical tone automaton. It is also common to have special constraints associated with final states, either raising for final high tones or lowering for final low tones, as shown for Tem, a Gur language spoken in Togo, case (c) in Figure 3. This process is trivially accommodated in the automaton by removing the final property from the low demiterrace state, and adding transitions to a separate final low state. Baule, a Kwa language spoken in Ivory Coast, has more complex contexts than most languages for tone sandhi; these can be formulated (non-deterministically) by adding additional oscillations; cf. case (d) in Figure 3.

Ega, an endangered language spoken in southern central Ivory Coast, is a three-tone language and, like a number of other languages with more than two tones does not have terracing. Tonal sequences in Ega tend to create a perceptual impression of chanting, rather than of descending terraces. An Ega example illustrating the chant-like tone alternation between two levels in the medial portion of the utterance is given in Figure 4; the first word shows a different tone, and the last shows final lowering and low tones (the very high and very low excursions are due to segmental perturbations). The mid tone stays close to 125 Hz, the high tone to 150 Hz, in musical terms a difference of about 3 semitones, i.e. a minor third (an interval which recurs again and again in the description of ‘stylised’ pitch patterning; whether

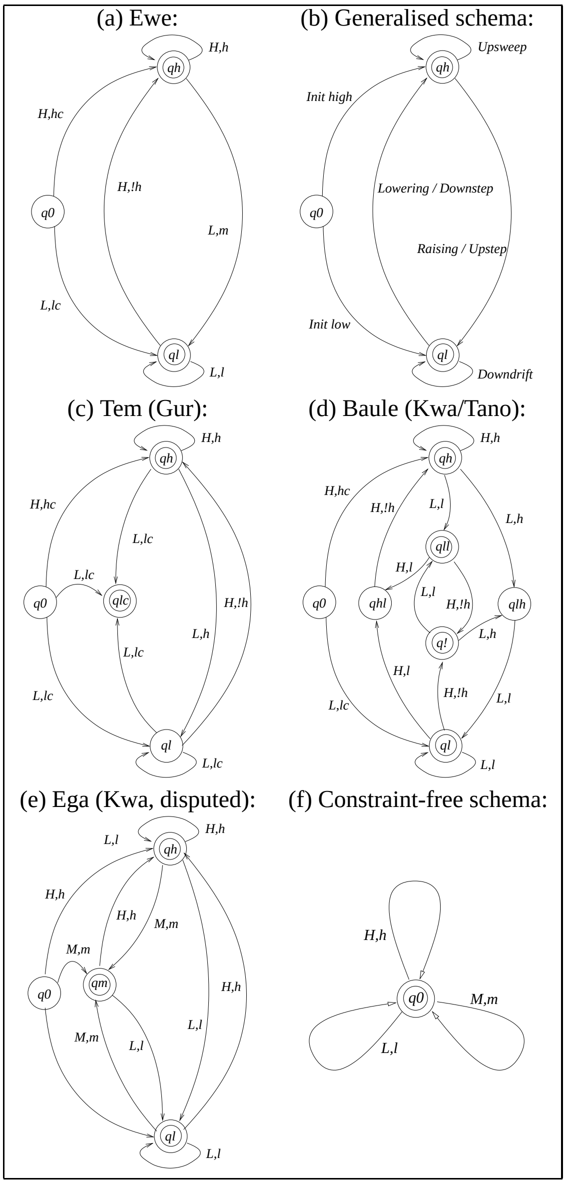

Figure 3: Typologically distinct lexical tone automata.

this is a significant constant is a much debated empirical question which cannot be addressed here).

The additional tone can be accommodated easily by adding another state, case (e) in Figure 3. This is justified because there are other constraints in addition to terracing in Ega, but if there were no such constraints, in fact the system could be collapsed to a single state with just one transition for each tone, as in case (f) in Figure 3.

4. Corpus processing with lexical tone

Currently two applications for the FST modelling method are being developed. Both start with a segmental transcription time-aligned with the recorded speech signal, as shown in Figure 4. The first is the automatic acquisition of an audio concordance based on the time-aligned transcriptions. The semi-automatically constructed prototype has been described in [8], and work is continuing on enhancements with bidirectional ap-

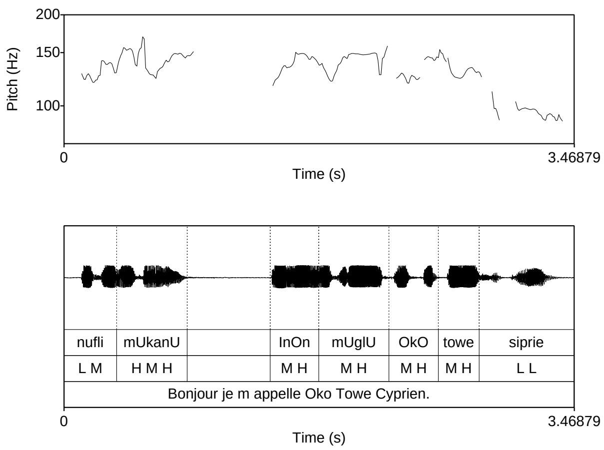

Figure 4: Chant-like discrete level pitch in Ega.

plications to (a) prediction of allotones with the FST model and assigning these to the corpus, and (b) prediction of lexical tones from allotones which have been manually assigned directly to the corpus.

The second application is the exact phonetic modelling of the fundamental frequency corresponding to tone sequences: the FST generates allotone sequences (e.g. with h, $!h$, l, m, lc ...), each of which is associated with an F0 transition function. An exhaustive search pattern matcher program iterates through values of the parameters of these functions. Typical measured and predicted values for a female speaker of Ibibio (a two-tone Lower Cross language spoken in Nigeria [9]) are shown below. Each utterance was assigned a tone sequence template generated by an Ibibio FST, e.g. t1, t2, t3, t4, t5 and t6 with t1 being an initial H, t2 a H, t3 a downstepped H, t4 a raised L between two Hs, t5 a L and t6 a final L. Search for the best fit produced results of the following type:

```
Data:         246.3 266.85 226.91 222.84 199.3 198.57
Model:        246   266    227    225    197   196
Pattern code: t1    t2     t3     t2     t3    t2
Allotones:    H/#_  H      !H     H      L/H_H H
```

In this example, the model encodes different allotones: initial H, (`H/#_`), downstepped H (`!H`), H elsewhere, L between two instances of H (`L/H_H`). Detailed results are given in [10]. Related studies on tone and intonation, using different pattern matching methods, are described by [11], [12], [13], [14]. Tests on interfacing the metrical FST approach with these other approaches to phonetic mapping are in progress.

5. Results and open questions

It has been shown that prosodic corpus processing for African tone languages can be given a well-defined formal and descriptive foundation. Further, it has been shown that the automata which are used to model formal linguistic descriptions can be used for practical purposes in audio concordancing, and for experimental purposes in fundamental frequency pattern matching. The use of a finite state transducer system ensures flexibility in applications, by providing a bidirectional mapping between lexical representations of tone and the allotonic representations which may be expected in a corpus with non-theoretically directed prosodic markup.

Interesting issues beyond the scope of the present study include segmental reduction and the generation of floating tones in fast speech, the automatic detection of the most appropriate model for a given language, and the extension of finite state modelling to other aspects of African languages. It is intended to use the approach in language planning contexts in cooperation with the University of Cocody, Abidjan, and the Ivory Coast authorities, in the production of language materials, including a component of an experimental speech synthesis system for Jula (Dioula), Mande, one of the major West African vehicular languages.

6. References

[1] Ahoua, Firmin (1996). *Prosodic Aspects of Baule With special reference to the German of Baule speakers*. Cologne: Köppe Verlag.

[2] Carson-Berndsen, Julie (1998). *Time Map Phonology: Finite State Models and Event Logics in Speech Recognition*. Dordrecht: Kluwer.

[3] Pierrehumbert, Janet (1980). *The Phonetics and Phonology of Intonation*. Diss. MIT.

[4] 't Hart, Johan, René Collier & Antonie Cohen (1990). *A perceptual study of intonation: An experimental-phonetic approach to speech melody*. Cambridge: Cambridge University Press.

[5] Gibbon, Dafydd (1987). Finite State modelling of tone languages. COLING 87. Proceedings of the Conference on Computational Linguistics. Copenhagen.

[6] Goldsmith, John (1990). *Autosegmental and Metrical Phonology*.

[7] Kaplan, Ronald M. & Martin Kay (1994). Regular models of phonological rule systems. In *Computational Linguistics* 20: 331-378.

[8] Gibbon, Dafydd, Nils Jahn, Soma Ouattara & Thorsten Trippel (2001). Preliminary Specification, Design and Proof-of-Concept Implementation of a Portable Audio Concordance (PAC) RFC 1.0. Universität Bielefeld, DOBES Technical Report 4.

[9] Urua, Eno-Abasi (1996/97). A phonetic analysis of Ibibio tones: a preliminary investigation. *Journal of West African Languages* XXVI.1: 15-26.

[10] Gibbon, Dafydd, Eno-Abasi Urua & Ulrike Gut (2000). How low is floating low tone in Ibibio? 30th Conference on African Languages & Linguistics, U Leiden, 27-30 Aug 2000.

[11] Connell, Bruce A. & D. Robert Ladd (1990). Aspects of Pitch Realization in Yoruba. *Phonology*, 7.1, 1-29.

[12] Bird, Steven (1994). Automated Tone Transcription. *Proceedings of the First Meeting of the ACL Special Interest Group in Computational Phonology*, 1-12, Las Cruces.

[13] Laniran, Y. (1992). Phonetic aspects of tone realisation in Igbo. Progress Reports from Oxford Phonetics 5: 35-51.

[14] Liberman, Mark, J. Michael Schultz, Soonhyun Hong and Vincent Okeke (1993). The phonetic interpretation of Igbo tones. *Phonetica* 50: 147-160.