Nuclear stress in eastern Benue–Kwa (Niger–Congo)

Update 2 November 2010: Encouraging agreement with this analysis in many empirical as well as theoretical respects (though with significantly different formal implementation) can now be consulted in the following publication:

Abstract

In 'Bantu' (eastern Benue-Kwa), standard analyses of information structure assume autosegmental tonemes and templatic verb morphology (Odden 1984; Hyman 1999). Dispensing with both devices, intonation tones reduce to phrasal accents (Bamba 1991; Idsardi & Purnell 1997), a VP constituent emerges from aux and clitic shells (Keach 1986; Myers 1998; Kinyalojo 2003a,b) and focus prosody appears Romance-like, with topics dislocated away from nuclear stress (Vallduví 1990; Zubizarreta 1998). The revised view enhances learnability, narrowing the set of possible languages while capturing language-internal generalizations.

1. Caveat stressor

In generative studies of Germanic and Romance, the term nuclear stress refers to default phrasal accentuation which codes information focus (Chomsky 1971, 199f.; Vallduví 1990; Zubizarreta 1998). The concept isn’t universally applied, thanks to a stack of gratuitous assumptions. (i) Prosodic phonology (Pierrehumbert 1980; Selkirk 1995) views accent as “non-phonemic prominence structure having to do with various pragmatic things people do” and thus as an unlikely source for “symbolic representations of phonological units that determine pitch differences” (D. Odden, p.c.). In English, however, nuclear stress is phonologically conditioned (Chomsky & Halle 1968, 17; Arregi 2002) and also shows “genuine ambiguity” between broad and narrow focus interpretations (Ladd 1996, 201; cf. Gussenhoven 1983b), therefore accent can’t be excluded from derivations and a null theory of accent is the null hypothesis (Cinque 1993). (ii) Cartographic syntax (Brody 1990; Rizzi 1997) treats ex-situ focus as attraction to a formal/“criterial” feature with no intrinsic PF correlate. In Hungarian, however, subordinating main stress (the “Eradicaton Rule”) tracks focus movement (Kornai & Kálmán 1988; Szendröi 2003). (iii) Relativistic descriptions of prosody (§1.1) and morphology (§1.2) underanalyze focus phenomena in Kimatuumbi (§2) and other ‘Bantu’ languages (§3). ¹

1.1 Autosegmental hallmarking

Tonemes—alphabetic representations of paradigmatic pitch contrasts (Pike 1948)—were ported into generative grammar as unscreened conceptual baggage three decades ago (Williams 1971; Leven 1973; Goldsmith 1976). Less than robustly detected in instrumental and perceptual experiments (Rouget 1965; Léveillé 1974; Hombert 1976; Abramson 1979; Gandour 1983; Connell 2000; Dilley 2005), tonemes have proven inseparable from metrical effects like interpolation, skewed distribution and juncture (McCawley 1978; Abramson 1978, Akinlabi 1985; Sietsma 1989; Hagaruchi 1991; Bamba 1991; Lázník 1992; Liberman & al. 1993; Liberman 1995; Idsardi & Purnell 1997). Despite such clues, tonemes are immune to disproof if accents are considered as just another tier in an “autosegmental-metrical” array (Leben 1976; Halle & Vergnaud 1982; Goldsmith 1982; Beckman & Pierrehumbert 1986; Gussenhoven 2004).

The main justification for tonemic autonomy is that phonological contrasts obey phonetic biuniqueness, as in Hyman’s slogan “tone = pitch features present lexically” (1989, 116) or in Odden’s sorting PF properties into “stress” versus “tone” while rejecting “shared formal machinery” despite “certain apparent similarities” (1999, 212). But biuniqueness loses generalizations in taxonomic discovery procedures (Halle 1959, 23), translation wordlists don’t reliably find minimal pairs, and language-internal puzzles result. (i) Emic tone demands ad hoc junctures. In Kimatuumbi, for example, Odden observes “tonal upsweep”—a gradual pitch rise from “L” to “H” across multi-syllables and a phrase boundary— and writes it as a string of H tonemes separated by [¡], the mark of “phonetic upstepping” (Pike & Wistrand 1974; Meir 1975; Vergnaud 1982; Goldsmith 1982; Beckman & Pierrehumbert 1986; Gussenhoven 2004).

In generative studies of Germanic and Romance, the term nuclear stress refers to default phrasal accentuation which codes information focus (Chomsky 1971, 199f.; Vallduví 1990; Zubizarreta 1998). The concept isn’t universally applied, thanks to a stack of gratuitous assumptions. (i) Prosodic phonology (Pierrehumbert 1980; Selkirk 1995) views accent as “non-phonemic prominence structure having to do with various pragmatic things people do” and thus as an unlikely source for “symbolic representations of phonological units that determine pitch differences” (D. Odden, p.c.). In English, however, nuclear stress is phonologically conditioned (Chomsky & Halle 1968, 17; Arregi 2002) and also shows “genuine ambiguity” between broad and narrow focus interpretations (Ladd 1996, 201; cf. Gussenhoven 1983b), therefore accent can’t be excluded from derivations and a null theory of accent is the null hypothesis (Cinque 1993). (ii) Cartographic syntax (Brody 1990; Rizzi 1997) treats ex-situ focus as attraction to a formal/“criterial” feature with no intrinsic PF correlate. In Hungarian, however, subordinating main stress (the “Eradicaton Rule”) tracks focus movement (Kornai & Kálmán 1988; Szendröi 2003). (iii) Relativistic descriptions of prosody (§1.1) and morphology (§1.2) underanalyze focus phenomena in Kimatuumbi (§2) and other ‘Bantu’ languages (§3). ¹

1.1 Autosegmental hallmarking

Tonemes—alphabetic representations of paradigmatic pitch contrasts (Pike 1948)—were ported into generative grammar as unscreened conceptual baggage three decades ago (Williams 1971; Leven 1973; Goldsmith 1976). Less than robustly detected in instrumental and perceptual experiments (Rouget 1965; Léveillé 1974; Hombert 1976; Abramson 1979; Gandour 1983; Connell 2000; Dilley 2005), tonemes have proven inseparable from metrical effects like interpolation, skewed distribution and juncture (McCawley 1978; Abramson 1978, Akinlabi 1985; Sietsma 1989; Hagaruchi 1991; Bamba 1991; Lázník 1992; Liberman & al. 1993; Liberman 1995; Idsardi & Purnell 1997). Despite such clues, tonemes are immune to disproof if accents are considered as just another tier in an “autosegmental-metrical” array (Leben 1976; Halle & Vergnaud 1982; Goldsmith 1982; Beckman & Pierrehumbert 1986; Gussenhoven 2004).

The main justification for tonemic autonomy is that phonological contrasts obey phonetic biuniqueness, as in Hyman’s slogan “tone = pitch features present lexically” (1989, 116) or in Odden’s sorting PF properties into “stress” versus “tone” while rejecting “shared formal machinery” despite “certain apparent similarities” (1999, 212). But biuniqueness loses generalizations in taxonomic discovery procedures (Halle 1959, 23), translation wordlists don’t reliably find minimal pairs, and language-internal puzzles result. (i) Emic tone demands ad hoc junctures. In Kimatuumbi, for example, Odden observes “tonal upsweep”—a gradual pitch rise from “L” to “H” across multi-syllables and a phrase boundary— and writes it as a string of H tonemes separated by [¡], the mark of “phonetic upstepping” (Pike & Wistrand 1974; Mair & al. 1975). But only the final H in such a sequence is at the pitch level typical for an H-tone (Odden 1996, 6), so to specify each middle syllable as H entails an absurd rule of

...
Duanmu (1994, 567; 2005)

(1a) “... stressless syllables either do not carry underlying tones... or will lose their underlying tones”

(1b) “... while languages can differ in word stress, all languages have the same rule for compound and phrasal stress”

Wagner (2005, 34, 273, 275) 3

(2a) “If A and B are sisters and A is the functor and B its argument, ... B is [accentually] more prominent than A unless A already contains an argument...” (cf. Cinque 1993, 244).

(2b) “Only associative domains are built in a single cycle. ... Each cycle consists of a right-branching structure.”

(2c) [An informationally given constituent] “becomes a functor” (cf. Steedman 2000).

Hyman (1999, 153)

(3a) “[– focus] → tonal integration = reduction”

(3b) “[+ focus] → tonal finality = end demarcation”

Despite the morphosyntactic distance between these languages, patterns (1) - (3) are empirically nondistinct. This convergence could be due to extrinsic factors (4a), or it could show that the faculté de langage handles stress and tone uniformly (4b).

(4a) prosodic relativity: focus cues are incommensurable grammaticalized “strategies” (Vallduví & Engdahl 1996), interpreted beyond the semantic interface (Reinhart 1997) under a generic functional label of “prominence” (Truckenbrodt 1995).

(4b) prosodic unity: all natural languages compute semantically-relevant prosody as phrasal accent (‘nuclear stress’). (4a) is unfalsifiable (cf. Kaye 1988), both because it’s evaluated in a global output procedure (economy/OT) and because it admits any imaginable focus cue (pitch, duration, word order, morphological Gestalt...). 4 To limit indeterminacy, it’s sometimes suggested that a focus-sensitive feature can’t also support lexical contrasts—e.g. “… the phonetic cue to sentence accent is duration in Bantu languages” since “Bantu languages cannot use the prevalent component of pitch changes because of their tonal nature” (Zerbian 2005, 15)—but such an inference is false in general. In East Asian languages, lexical tone doesn’t block “parallel encoding” of focus information as F0 pitch (Xu & Xu 2005; cf. Potisuk & al 1996; Xu 2004), and in ‘nontonal’ Chimwini-Kiswahili, vowel duration is lexically contrastive as well as governed by focus-sensitive phrasing (Kisseberth 2002). Appeals to lexical tone are anyway circular: it’s “difficult to draw a dividing line between languages with contrastive tone on (almost) all syllables and languages with tone contrasts in more restricted locations in the word. Standard Chinese and Swedish are... both tone languages by this definition” (Gussenhoven 2004, 47). Eastern BK is equivocal in these terms, because lexical pitch is underspecified in all conceivable ways: paradigmatic (Meeussen 1963; Stevick 1969), syntagmatic (Voorhoeve 1973) and relative to syntactic category (Odden 1988; Kimenyi 2002). The literature does report one plausible instance where duration necessarily substitutes for pitch as a focus cue: in second-occurrence focus—however this effect depends on configurational anaphora not lexical contrast sets, and implicates metrical structure (Rooth 1996, 219) consistent with (2c).

(4b) on the other hand could be wrong, but finds support beyond the studies cited under (1) - (3). In an elegant analysis of Spanish and Catalan, for example, verbs and nouns follow different stress rules reflecting the difference between high and low pronunciation of the root in its extended phrasal shell (Arregi & Oltra-Massuet 2005). To be sure, it’s not obvious from inspection of BK citation forms that every (OCP) H tone recovers an accent, thus implying the presence of a local weak (non-H) position in keeping with metrics’ “relational” property (Liberman 1975, 51). It wasn’t even obvious in English, to English-speaking linguists, just 60 years ago: anticipating ToBI (Pierrehumbert 1980), Pike misses the phrasal distribution of stress (Bolinger 1951; Chomsky & al. 1956, 72) and sets up four “relative but significant levels (pitch phonemes)” (1945, 25), stringing them into intonational lines with a redundant word-level stress [‘] for each “2” tone. 5

(5) a. He wanted to do it. [‘2-4]  

b. I want to go home. [3-‘24]

c. The boy in the house is eating peanuts rapidly. [3-‘23-3-‘23-3-‘23-‘23-‘24]

In sum, the tonemic valise packs a load of warrantless assumptions plus a bundle of pitch features in search of analysis. In eastern BK, this search meets a second type of impediment.

1.2 The diacritical imperative

Some Bantuists identify a focus slot right-adjacent to the finite verb (Watters 1979; Ndayiragije 1998; v.d. Wal 2005), but this can’t be the verb’s syntactic complement— as in Hyman’s (1985b) analysis of Aghem—if one recognizes a polysynthetic “verbal word” (Nurse 2003, 90, cf. Meeussen 1967; Goldsmith 1985; Hyman 2003), a templatic string of argument-type clitics, an aux, the lexical root plus derivational
“extensions” (minimally, a default vowel), cf. (6a). Canonical in conjunctive orthography (Guthrie 1948), the template is reborn in generative theory (Baker 1988, 1996; Odden 1996, 71; 228f.; Carstens 2001) and entails computational explosion: Odden (1981, 17) reportedly estimates “that some 16,000,000,000,000 distinct forms can be built around a single verb radical, not counting the differences induced by distinctive tonal features of various morphemes” (Sietsema 1989, 90). Luckily, a range of evidence shows that the left edge of the verb root is separated by ordinary phrase boundaries from subject “agreement” (CLs), which is “merged as a DP, an independent syntactic object” (Kinyalolo 2003b; cf. Keach 1986; Bresnan & Mchombo 1987; Myers 1998; Russell 1999; Kinyalolo 2003a), cf. (6b).7

(6)a. \[ X° CLs–(aux)–(CLo)–verbroot–extensions \] \[ DP+foc … \]

(6)b. \[ TP [DP CLs] \] \[ T° (aux) [KP [DP (CLo)] [VP [V° verbroot–extensions] [DP+foc … ]]]]

The choice between (6a) and (6b) has a clear consequence for focus: (6b) allows a surface VP constituent, whereas under (6a) a VP must be simulated by stratally-ordered lexical phonology (Pulleyblank 1983; Odden 1996, 228ff.) plus a focus diacritic. This diacritic is actually embraced:

“Bantu languages in particular are known for their ‘focus prominence’ … as when a tense is marked differently according to whether the verb is included in the focus or not. … Whether one is a syntactician or semanticist wishing to study focus or one is a phonologist wishing to study tone, one must consider all aspects of the grammatical system of a Bantu language. … [T]o not do so would be to risk drawing the tempting—but wrong—conclusion that there is a direct link between semantic focus and pitch in these languages.” (Hyman 1999, 151, 174, his italics)

Similarly, Odden rescues (6a) with a semantic parameter, making Kimatuumbi a language where “focal-sensitivity is a general property of an entire grammar” banning “two items focused in a clause” and deploying “morphological processes whose sole purpose is marking focus”; the existence of examples like “Who likes only meat? or Tom likes Sally” are said to show that such a ban “is not found in languages like English” (Odden 1984, 277, his italics, cf. 1996, 71).

On second thought, however, these arguments for (6a) fail. Hyman’s rhetoric above is misdirected and overblown. Misdirected: it’s not the “syntactician or semanticist” but rather the prosodic phonologist (Pierrehumbert 1980; Gussenhoven 1983a) or interface economist (Reinhart 1997) who’s “tempted” to draw “a direct link between semantic focus and pitch”. Overblown: no reason is given why the road of indirectness must implicate “all aspects of the grammatical system of a Bantu language” (my italics), since just one grammatical feature (syntactic phrasing) may suffice. Odden’s English examples involve operator absorption not multiple independent foci (Krifka 1991), and an argument type expression modified by the Kimatuumbi item translating only, “whose sole purpose is marking focus” (cf. Rooth 1985), requires a “noun focal tense” (Odden 1984, 292). In any case, semantic parameters are doubtfully learnable (Gavarró & al 2005), so they’re slim support for exotic morphosyntax.8

The remaining task is to reconsider focus in eastern BK without tonemic and templatic distractions. Although I lack firsthand experience with these languages, the standard literature contains numerous helpful hints.

2. Kimatuumbi

Odden holds that “there are no data in Kimatuumbi which help choose among competing definitions of focus” (1984, 278), but empirical differences do distinguish the diacritic/tonal (§2.1) and syntactic/accentual (§2.2) analyses.

2.1 H deletion, diminishing returns

Odden generates Kimatuumbi verbal prosody by lexical H-tone assignment, followed by morphosyntactic tone mapping:

“Every verb is assigned a floating H-tone, which is mapped… to the third vowel after the subject prefix in the subjunctive and participial; it is mapped to the second mora in dependent clauses; otherwise it is mapped to the root-initial mora. …[T]o the extent that H is assigned to different moras in the stem, we have support for the independence of these processes” (Odden 1996, 191f.)

By this premise, every Kimatuumbi verb should end up with one H, but actual occurrences range from zero to two, and that’s where focus comes in. Sticking to main clause indicatives, verbforms are subclassified along a “three-way contrast in focal properties”: (7a/b) are “neutral”, (7c/d) “noun-focal” and (8a-d) “verb-focal” (Odden 1984, 289, 295 fn), although all the forms in (8) are systematically ambiguous between focus on VP as a whole, and narrow V. In accordance with the Bantuist template (6a), even the auxiliated, double-H forms in (8) count as “simple verbs” (Odden 1996, 71).9
In retrospect, these data are less friendly to morphological focus than advertised. The “focus-neutral” status of (7a/b)—illustrated in (9) - (11) below—betrays the notion that “focal-sensitivity is a general property of an entire grammar” (Odden 1984, 277). A three-valued focus feature overgenerates, providing no clue as to why Kimatuumbi lacks a “noun focal” past tense; why a “neutral” progressive specially requires periphrastic ‘be’; or why the “nine forms of the future [are] all focally neutral” (Odden 1996, 62f., my italics).10 Barring accident, these gaps show that focus is not a morphological feature; rather, focus interpretations emerge as a compositional product of freestanding elements. Even taking the morphology at face value, if natural language has at least one generative engine—a syntax—and conceding that morphology can emulate at least some syntactic effects (Keenan & Stabler 2003), then for any given interpretation—including focus—a Kimatuumbi child needs to decide whether syntax or morphology is driving it. Odden admits that the motor is non-morphological in certain “neutral” forms which yield narrowly verb-focal readings even though unassisted by dedicated “verb focal” inflection (1996, 62; 1984, 281):11

From (9b), Odden reasonably concludes that “the focal properties of the verb are determined independently and are not a direct result of the selection of the verbal morphology” (1984, 281f., fn. 4). Logically the next steps would be (i) to name these independent factors; and (ii) to check whether the same factors also explain “noun-focal” and “verb-focal” effects. Attaining step (ii) would mean that morphological focus diacritics perform no indispensable work in Kimatuumbi.

As to step (i), notice that both examples in (9) display object pro-drop. Assuming that this ellipsis is necessary in order to obtain narrow verb focus in “neutral” forms, it follows that focus is only accidentally narrow in (9)—just as expected for nuclear stress in a phonetically nonbranching VP.12 In the same vein, consider a wider sample of focal interpretations which can contextually elicit a “neutral” form (Odden 1984, 280f.):

With direct object in situ, the neutral form returns either broad VP focus (10a) or narrow focus on the object (10b)—a situation familiar from English. Narrow subject focus (10d/e) requires VS order, as in Italian (Antinucci & Cinque 1977, 124).13 Focus on the possessor of the complement (10f) is also consistent with nuclear stress, assuming that the postnominal possessor of Kimatuumbi is c-commanded by its possessum.

Not only does the “neutral” form have a semantics consistent with nuclear stress, it also has the phonetic earmarks. Apparently without exception, a main verb bears H in two circumstances: after an aux (7b, 8, 9b, 10) or phrase finally—even absent an aux (11e). But with no aux, any clausalmate phonetic material to the right of the verb, as in (11a-d), is sufficient to block H on the verb itself (Odden 1996, 62, 233, 287)—conforming to the first clause of (2a) above.14
(11)a. A kalang-ître yöopáča čē. ‘S/he recently fried [something] to get money’
   b. A kañ-ître kāmílba. ‘S/he recently cut rope’
   c. A tel-ike námá ñū. ‘S/he recently cooked only meat’
   d. A tel-ike Mambóondo. ‘M. recently cooked [something]’
   e. Mambóondo a tel-ike. ‘M. recently cooked [something]’

As mentioned, Odden presyntactically assigns H to all verbforms, then maps this H to the left edge of the root in finite main clauses, and lastly deletes the same H in case the verb is not clause-final (7a, 9a, 11a-d) by “Perfective Tone Loss” (1996, 233). Curiously, though, a deletion sensitive to the righthand context happens to fail just in case the root is introduced by an overt aux—a lefthand context exception which accommodates the tonal contrast between the “recent” (7a) and “remote” (7b) perfectives. The (7b) form shows that the absence of root H depends on a larger domain than [VO]. (7b) aside, the more general problem with the sequence of lexical H insertion followed by phrasal H deletion is that it is theory-internal: there can be no independent evidence for the deletion, because Kimatumbi happens to be a language “lacking lexical tones in verbs, ... similar to Safwa, Kinga, Makua and Kikuria” (Odden 1996, 165; cf. Cheng & Kissberth 1979, 31). The opposite conclusion—that the H in (7b, 11e) is epenthetic and phrasal, and therefore accentual in the normal meaning of the term—is compelled by an additional fact: H also shows up in unauxiliated progressives (forms which are not “perfectives”) when these are preceded by an object clitic (Odden 1996, 192), thus the H-less forms in (7c/d) contrast minimally with (12a-b): 15

(11)a. *tãlek-ñúte kalang.* ‘... is cooking sweet potato(es) for her/him’ ['her/him' = n]
   b. *tãlek-núte kalang.* ‘... was cooking sweet potato(es) for me’ ['me' = nj ]

To handle (7c/d) as well as (12), Odden suggests to “postulate a special rule deleting H from a mora that is both stem initial and superstem initial... in noun focal tenses” (1996, 195). 16 This deletion won’t occur in (12) because the H in (12) is not superstem-initial—Odden’s “superstem” being a sub-word constituent which includes the object clitic. Of course this “special rule” is a mere restatement of the problem, but worse it is inadequate due to the presence of the object clitic between the subject clitic and the root in both examples of (13), H deletion should fail to apply, but no verbform H is reported either in applicative (13a), in which direct object scrambling puts default focus on the applied object, or in causative (13b), where the “superstem”-initial clitic receives a causee interpretation (Odden 1984, 74, 296): 17

(13)a. *Námá niñ n kalang-ý-a Mambóondo.* ‘I’m frying meat for M.’ ['M.' = n]
   b. *Tũy n kalang-ý-a Lübũlũlee Námá.* ‘We’re making L. fry meat.’ ['L.' = n]

The dim prospect of fiddling with “special” H-deletion so as to affect (13) without wrongly also applying in (12) shows that tonemic investment yields diminishing returns: morphological, focus-sensitive tone rules lack generality because tone deletion lacks a coherent structural description, needing a new patchup rule for every new context.

2.2 Clitics all the way up

If tonemes are the lexical hypostasis of accent (§1.1 above), a natural alternative to H-deletion as an account of focus prosody is to turn the toneme ‘on its head’ and express the complement set of H-deletion contexts as accent-driven H-epenthesis. Doing this for Kimatumbi, given the absence of lexical pitch contrasts on verbroots, leaves syntactic phrasing as the sole basis for verb accentuation in the language. As already discussed (§1.2), this is blocked by the morphological template (6a) as assumed by Odden (1996, 71, 228f), but it’s possible with the phrasal syntax of (6b) which admits a surface VP and so brings focus effects in reach of nuclear stress as in (2). Tentative results can be glimpsed in diverse constructions.

Unauxiliated progressives being “noun-focal tenses have no prepausal form in main clauses” (Odden 1996, 195), thus the nuclear stress rule (2a) can’t explain H-epenthesis in these forms—nor should it, if extrinsic factors are responsible. 18 One such factor, seen in (12) above, is where stacking an object clitic between the subject clitic and the root evokes a lexically spurious H. Assume with Seidl (2001) that the object clitic is licensed in a KP shell whose null head doubles thematic material in VP: the applicative formative [j] ~ [y]. 19 Then the cross-root dependency between KP shell and applicative extension, flagged in (12) by coindexing, diagnoses a spellout domain as in (2b): the lexical predicate can’t be construed with two structurally external arguments symmetrically, therefore KP is accentted. This accent is however realized neither on the object clitic, nor on the null head of KP, but on the nearest accentable syllable, which happens to be on the left edge of VP. Auxiliated, nonapplicative (7b) is parallel: the verbroot is linearized between the overt aux [a] and the aspectual suffix [ite]; regardless of whether this suffix is an...
argument-type expression (Manfredi 2005b) or a predicate operator, its functional composition within TP is nonassociative, therefore by (2b) Asp inhabits a different cycle from the subject/tense, so AspP is accented. Completing the parallel to the KP shell in (12), the AspP shell in (7 b) is phonetically sub-minimal—in this case null—and again accent is realized on the nearest accentable material, namely the left edge of VP.20

The presence of VP-initial accent in (12) contrasts with its absence in (13), though the object clitic occurs throughout. This difference is beyond the expressive power of tonemic-templatic description, as pointed out above: no revision of the H-deletion rule can cover both (12) and (13) at once. Considered in phrasal terms, however, a differential effect does appear: for construction-particular reasons affecting the position of the direct object, both (13a) and (13b) display a discontinuous VP constituent, and the relevance of this fact to accentuation is consistent with (2a).21

Another lexically spurious, phrasally-assigned H appears at the right edge of any XP before a conjoint/adjoint phrase or clause. The rule’s name, “Phrasal Tone Insertion” (Odden 1996, 234-38), tells the whole story:22

A third type is the “verb-focal” paradigm, cf. (8) above. The phonetic generalization is that the inflected domain (aux plus verb) displays two pitch peaks (call them H1 and H2) whose distributions are patently phrasal. H1 occurs at the right edge of the (minimally CV) aux or aux stack.23 H2 is found at the right edge of the lexical V° (extension included). The tonemic-templatic approach treats H2 as derivationally identical to the H that shows up immediately after the aux in “focally neutral” (7b); by Odden’s rule of Focal Flop, the first H gets “delinked” and “set afloat — later… docking… to the final syllable” of the verb (1996, 193). Viewed accentually, however, the root-initial H in (7b) is only superficially linearized on the verb, but is generated on a later cycle. As to H2 in (8), its own appearance depends on the appearance of H1, moreover its position apparently at the right edge of VP draws semantic support from the fact in (8) that the object cannot bear narrow scope: what is in focus in (8) is either the verb by itself, or the entire V+O sequence if any. It’s incorrect to describe the interpretation of (8) as “translated into English with contrastive stress on the verb” (Odden 1984, 279). That may be true in (15a), necessarily read with narrow scope on the verb, and it may indeed be “preferable to topicalise an object noun phrase when a verb-focal tense is selected” (Odden 1984, 296), but (15b/c) with the object not topicalized are both cited as felicitous answers to questions with broad VP scope (1984, 280, 290). A narrowing of <VP> focus to <V> in examples like (15a) could be imposed contextually, easier than a widening from <V> to <VP>.24

The hard question for accent is how the position of H at the right edge of V° tracks VP focus. It must be the case that the object is evacuated from VP without removing it from the focus domain, but this surprising inference is supported by several considerations. Phonetically, Odden wonders why “the final H-tone which derives via Focal Flop cannot undergo Retraction” onto a preceding double vowel (1996, 199), as expected in the tonemic framework, yielding something like the ungrammatical [*[… kalaăng- aa… ]* in (8) or (15b). This failure of retraction banal, under the accentual assumption that the verb-final position of the H in (8) and (15) is already retracted from the complement, as expected for nuclear stress. Indeed, Kimatumbi presents independent evidence for retraction in (8) and (15): the concomitant failure to shorten the vowel of the verb root, which normally “applies when some word follows within the VP (Odden 1996, 226). For example the root ‘fry’ appears as
(16)a. CLs kalāang- ite. ‘... recently fried [something]’
    b. CLs a kalāang- ite. ‘... fried [something]’

(17)a. CLs tī kalāang- ā. ‘... recently fried [something]’
    b. CLs a tī kalāang- ā. ‘... fried [something]’

“It is not clear how this exceptionality is to be handled” says Odden (1996, 227) when shortening fails before an overt object in (8) and (15). But the syllabic and tonal facts follow if the object is phrased external to VP in those cases, as required by the accentual hypothesis. As to interpretation, more Kimatuumbi data are needed, but consolation comes from a remarkably similar phenomenon in Italian. According to Cardinaletti, the logical object in (18) is deaccented in-situ, no less than in (19) where it is patently dislocated when doubled by a clitic. In both contexts there is a “low pitch intonation contour... separated from the clause by an intonation break (signalled by a comma)” (2002, 30 fn 1). Despite the similar contour, Cardinaletti argues for a structural difference between (18) and (19). In (19b), for example, the participle is XP-extractable by itself, but this extraction fails in (18b). Conversely, (18c) but not (19c) allows a quantified direct object (Cardinaletti 2003, 36ff.).

(18)a. Ho già comprato, il giornale. (19)a. L’ho già comprato, il giornale.
   AUX 3s already bought       the newspaper
   ‘I’ve already bought, the newspaper’
   b. *Comprato, non ho, il giornale. b. Comprato, non l’ho, il giornale.
      neg aux 3s bought             the newspaper
      ‘I haven’t bought, the newspaper’
   c. Non ha invitato Gianni, nessuno. c. *Non l’ha invitato Gianni, nessuno.
      neg aux 3s invited G. nobody
      ‘G. hasn’t invited anybody’
   d. *Comprato, non ho, il giornale. d. Comprato, non l’ho, il giornale.
      neg aux 3s bought             the newspaper
      ‘I haven’t bought, the newspaper’

In sum, both Kimatuumbi “verb focus” (8) and Italian emarginazione (18) display a right-peripheral direct object which is outside of nuclear stress but still within sentential nuclear scope—though it may be excluded from that scope by further operations including dislocation. §3 concludes this paper with some parallels elsewhere in eastern BK.

3. “Disjoint” forms

Unlike Kimatuumbi (§2), most of eastern BK shows accentual prelinking (‘lexical tone’) in verbroots to some degree, but this difference doesn’t undermine the accentual treatment of focus as presented above.

As illustrated in (20) from Byarushengo & al. (1976, 196, 199) and in (21) from Hyman (1999, 153, 155), word order in Luhaya correlates with the prosody of penultimate syllables. [HL] is found in the penults of certain argument type phrases (‘nouns’) just in case they’re sentence-final (20, 21a), narrowly focused (20b) or clitic-doubled after the verb (20c/d, 21c), otherwise the penult is simple [H] (20a, 21a/b). As for certain finite verbs, the penult is L (20a/b, 21a) unless an object clitic precedes, in which case the penult is [H] if the verb is in narrow focus (20c/d) or is followed by a nondoubled argument (21a), otherwise the penult is HL (21b/c). To highlight these alternations, I use a simplified transcription exploiting tone/length redundancy, writing [xx] for HL, [x] for H and no brackets for no H.

(20)a. Abak[a]zi ba bon’ omw[aa]na. women   CL see child
   ‘The women see the child’ [broad focus]
      CL see child      women
      ‘They see the child, the women’ [narrow focus on the direct object]
      CL CL see   women child
      ‘They see him, the women, the child’ [narrow focus on the verb]
      CL CL see   child women
      ‘They see him, the child, the women’ [narrow focus on the verb]
   'The women are counting the goats for the child' [broad focus]

   'The women, the child, the goats, they are are counting them for him'

   'The women, the child, the goats, they are are counting them for him'

   'They are counting them for him, the women, the child, the goats'

Hyman maps H tonemes to surface distributions by autosegmental spreading and deletion, filling in L tonemes. A rule changing penult [H] to [H] is triggered before %, a diacritic that “marks assertive focus, i.e. the end of an assertion” (1999, 154). This entails that examples (21a/b), each with a single [H] contour, count as simplex assertions, but that (21c) with a total of four such contours contains four % diacritics and therefore “four completed assertions” in one sentence (1999, 156). While saving the diacritic approach to focus (§1.2) and the tonemic hypothesis (§1.1), Hyman also tacitly rejects a treatment of Luhaya % as a clause boundary, i.e. a formal object with only indirect, syntactically-mediated impact on focus interpretation. This nondiacritic view assimilates a Luhaya ‘noun’ after a % to a clitic “right dislocation” in Romance languages i.e. as a strategy that “allows speakers to remove an argument, or other bit of information from within the scope of assertion” (Byarushengo & al. 1976, 197, crediting Francesco Antinucci p.c.). The same view, according to which “the effect of the % boundary is to ‘defocus’ non-asserted information” (Byarushengo & al. 1976, 198), is close to the syntactic treatment of Kimatuumbi in §2.2 above, modulo the extra option of prelinked H in certain verbroots as well as the richer, triple contrast of syllable prosodies (HL, H, ∅). Both considerations favor the accentual treatment of H: a phrasal regime of nuclear stress expects the nonparsing of H in an unperturbed, broad focus VP lacking object clitics (20a/b), whereas a narrow focus shift can be blamed on use of a focus diacritic which has at most two values, minus (3a) and plus (3b). An indirect analysis which reads focus from syntax naturally admits more than one kind of focus related phrasing.

Based on available published data, “it’s hard to argue directly against Hyman’s (1999) theory-internal claim of “four completed assertions” in (21c). A sharper problem for the claim is posed by narrow focus data like (20b-d): the most natural interpretation of the English glosses allows one assertion per example, despite the presence of multiple [H] contours, so either the glosses, or the multiple focus analysis, must be wrong. Conversely, the freedom of post-verb word order in (20c/d) can be taken as evidence for the nondiacritic 1976 treatment of %—as a sentence boundary lacking intrinsic semantic content—in view of the parallel word order freedom in Spanish examples like (23), which has been argued to diagnose right dislocation (Zubizarreta 1998, 156).32

Another reason not to build semantic assertion into the % boundary is that real semantic assertion is transparently implicated in a different Luhaya contrast, discussed by Hyman (1999, 160-62) under the rubric of “conjoint/disjoint” phrasing of post-verb material (cf. Schadeberg 2004). A relevant pattern in both languages is that lexically prelinked H is “suppressed” in environments of what Hyman & Watters (1984) call “auxiliary focus”. Because this distribution differs from the suppression of verbroot H in (20) and (21), conditioned by % and tied by hypothesis to a focus diacritic, Hyman is obliged to introduce “a secondary focus that has become morphologized and which, therefore, only imperfectly corresponds to the semantics that motivates it” (1999, 162). Assuming with Hyman that both phenomena are indeed focus-related, his dilemma can be blamed on use of a focus diacritic which has at most two values, minus (3a) and plus (3b). An indirect analysis which reads focus from syntax naturally admits more than one kind of focus related phrasing.

In Kirundi, as first observed by Meeussen (1959, 119-28), the “disjoint” aux ra is necessary for the appearance of H in the lexically accentuated root ‘pick’ in (23a). Absence of ra suppresses the H (23b). The prosodic contrast is missing with a lexically unaccented root, as in (24) from Ndayiragije (1998), leaving auxiliation as the only cue. The correlation of auxiliation and lexical H in (23) follows from nuclear stress if the aux and VP form separate cycles, as is guaranteed by the pleonastic character of ra, which would otherwise be blocked by economy. As to interpretation, the presence of ra allows either the verb root alone, or the whole VP, to constitute new information (24a), whereas a narrow focus shift on the object, as in a content question (24b), entails ra’s absence. The correlation is 100%, because the ra auxiliary is pleonastic apart from focus considerations, much like affirmative (i.e. stressed, declarative) English do, as helpfully hinted by Ndayiragije’s translation of (24a).
According to Goldsmith, Kirundi's "-ra- Focus marker... is itself a recent innovation shared with Kinyarwanda but no other languages (a reanalysis of what was formerly the present tense marker, presumably)" (1985, 127). Even so, the focus effect of auxiliation is not limited to this pleonastic form, as shown by Setswana. Literal pause is not necessary after a disjoint verb (Chebanne & al. 1997, 56), but disjoint phonetic traits remain prosodic in the broad sense: a language-particular mix of pitch/timing and affixation/auxiliation, both consistent with an analysis by phrasing. Créissels (1996, 110f.) reports systematic minimal pairs in the Setswana present perfect (25) and negative nonpast (26). In the superimposed paradigms in (27), the tonal difference appears in tandem with auxiliation, whether the aux in question is pleonastic (affirmative nonpast) or a contentful sentence operator (future). If the aux is substantive, the audible conjoint/disjoint distinction reduces to tone alone. In contexts with a nonroot character, including those labeled pluperfect or consecutive, the distinction is phonetically neutralized altogether, yielding radical ambiguity, (28).

Setswana differs from Kirundi semantically, in that the Setswana disjoint form puts narrow focus on the verb excluding the direct object (Créissels 1996, 113f.). A similar difference even divides Kirundi from Kinyarwanda, where the ra form excludes an adverbial (29a) or a direct object (30a) from the domain of new information (Givón 1975b, 194; tone outside the aux not marked in the source); and where the ra form is also impossible in a negative or relative predicate (reported but not illustrated in the source). The non-root character (30a) from the domain of new information (Givón 1975b, 194; tone outside the aux not marked in the source); and where the ra form can't be followed by a discourse-old (pronominalized or scrambled) object (31b), which the ra form allows (31a).

In Setswana, auxiliation is neither necessary nor sufficient for a disjoint form, which is narrow: (25a) is disjoint with no aux, whereas the nonfuture variant of (27a) is indeed auxiliated, but could also be conjoint depending on the tone (27b). In Kirundi by contrast, all disjoint forms described by Meeussen are auxiliated, with the further prosodic correlate noted above, and they are all broad. Crucially, in the Kirundi conjoint, the verb root is deaccented —a robust correlate of old information (Williams 1997)— so the fact that it's narrow is no surprise. Now, why is the Setswana conjoint broad? Prosody distinguishes two subcases. Assuming that accent appears as a [H] pitch contour, an accented verb is broad conjoint after a "middle field" aux—one which follows the subject clitic, as in (27b)—but narrow disjoint otherwise—if nothing separates the subject clitic and the verb root, as in (25), (26) and (28). In other words, focus projects from an accented verb iff an aux directly precedes. This generalization also holds in Kirundi, in fact trivially so because verb root accent (in this language, a simple H) is limited to auxiliated forms. Turning to Setswana sentences which lack a middlefield aux, an accented verb is narrow does not project, but this is not a comparative problem because Kirundi ra has no Setswana counterpart, for the independent reason already stated. The remaining case is an
unstressed (and therefore nonauxiliated) verb in Kirundi, which is narrow conjoint; the corresponding form in Setswana is broad. Both languages share this principle of stress-to focus mapping: focus projects from a stressed verb to VP (is broad) only after a Mittelfeld aux. In the absence of either VP-internal stress or an aux, focus includes the object (is conjoint) in both languages. Thusfar unexplained is the difference that neutral phrasing is broad in Setswana but narrow in Kirundi; tentatively this could be related to the different accentuation of argument type phrases (‘nouns’), presumably an independent fact.

4. Perspective
The foregoing shows that the choice to forego tonemes and Bantuist templates brings many BK prosodic phenomena under a general theory of phrase structure, a hypothesis of prosodic unity (4b), as well as a modular architecture of information structure (2c). Numerous language-particular consequences across BK and more widely remain to be drawn, e.g. the appearance of left-peripheral focus and “compulas” i.e. copular focus markers (Manfredi 1987, 110, cf. Bergvall 1987).
1. Theorit Bantu is sharp ideology but fuzzy linguistics. In apartheid Zuid Afrika it was the state synonym for “black”—as in this parliamentary speech by Minister P.W. Botha in 1964 (quoted by McGregor 2006):

“I am one of those who believe that there is no permanent home for even a section of the Bantu in the white area of South Africa and the destiny of South Africa depends on this essential point. If the principle of permanent residence for the black man in the area of the white is accepted then it is the beginning of the end of civilisation as we know it in this country.”

2. Odden (p.c.) notes that “[l]exicology isn’t the real issue, though that is often a fact that makes people comfortable with the idea of something being phonological.” Indeed, and once ‘tone’ is phrasal, a reanalysis in terms of nuclear stress comes within reach.

3. The “unless” clause of (2a), accentually weighting the predicate over the subject, can be dropped if the default focus operator is the middle field as in (2b).

4. Bamba & Liberman state that in Manding (Niger-Congo), “some of the functions of English intonational focus are performed by explicit and ordinary morphological marking” (1999, 1), but then go to note that this marking is homophonous with the non-negative copula—anything but “ordinary”!

5. The tonemic format of (5) may have a sociological explanation: Pike utilized the study of English intonation as one step in the teaching of [S.I.L.] students how to reduce tone languages to writing (1947, 131), but susceptibility to common sense Africanist analysis persisted half a century later, when for example Ladd defended “tonal targets” in English intonation based on “the existence of languages like Yoruba in which it is uncontroversial that the system of lexical tones is based on distinctive levels... Once we have such a theory, it is plausible to assume that it will apply to English or Dutch as well” (1996, 61).

6. Guthrie exempts from the template “some of the languages of the extreme north-western part of the Bantu area” (1948, 24 fn. 2). Templatists and even kinship have an affinity, e.g., in Yoruba (BK). Clark’s (1989) template, level ordering analysis—developing Goldsmith (1976) in tune with Clements & Goldsmith (1980)—is obliged to posit a phonemic “downstep” (an unpredictable tonemic juncture) whereas a syntactic, accentual analysis can derive this juncture from phrase boundaries, e.g., between aux and V or within the Genitive DP (Déchaine 1992; 1993, 497-520; Manfredi 1993; cf. Clark 1980).

7. Odden (1996, 228f.) considers the possibility that narrow focus on V results from verb raising, i.e. if the X 0 in (6a) = “Infl”, or alternatively from extraposition of [-focus] arguments. Neither kind of rephrasing is independently motivated, however. In (6b), the label V* is intended to cover the categorical confusions of “lexical syntax” à la Hale (1995).

8. I can’t find a Kimatumbi example with a ‘perfective’/focally ‘neutral’ verb and a noun phrase or PP modified by ‘only’. One may exist, given the situation with wh- expressions: Odden initially says that these “cannot appear with the verb focal tense, but may appear only with the noun focal tense” (1984, 292) but later observes that “Intrinsically focused elements such as wh-words may appear in clauses containing a focally neutral verb” (1996, 62). The crucial case for focus-sensitivity would combine a neutral verb, a nounclass. In examples (9) - (11) the subject clitic is [j], denoting a third person singular human subject which is traditionally called class 1 (Odden 1996, 34). Examples (7b, 8b, 9b, 10, 11) include an aux [j], labeled “remote” (Odden 1996, 57) and linearized immediately after the subject clitic. Thus (9b, 10, 11) show homophony between the default/epenthetic aux and third singular human pro. Welmers (1973) expresses scepticism about standard Bantuist tense labels denoting multiple degrees of remoteness.

9. A phrasal origin of focus in (9) is further supported by the doubled, specifically phrase-final form of negation (li-ŋ) in the second clause. A desideratum is defining the general freedom of object pro-drop in this language.

10. In contrast to Italian as well as to Kimatumbi, VS order denotes broad (“all new”) focus in Catalan (Vallduvi 1990), and VO is generally possible for narrow subject focus in Spanish (Zubizarreta 1998). Plausible sources of this variation—case assignment and clitic doubling—remain controversial in these languages.

11. The locution complement is used here rather than direct object, in order to denote an argument phrase in immediate post-verb position, thus including the focused, inverted subjects in (10d/e). This accords with Cinque’s view, that for purposes of nuclear stress, a non-topical post-verb argument is more deeply embedded than the verb itself “on the recursive side”. Nevertheless, the configurational definition is distinguished from an edge-based theory by cases of right detachment; faced with these, the edge-based theory needs to proliferate prosodic boundaries, while the syntactic view can refer to phrasal discontinuity.
15. For me, anaphoric describes minimal overt material not drawn from the numeration. For example, prosodic unity (4b) claims that children know how to construct metrical grids in nihilo, although of course they can learn lexicalized accents.

16. On the same page, Odden floats a second, more technical solution, without enthusiasm.

17. Both exceptional forms have two overt phrasal objects; I assume that scrambling of the direct object is obligatory in (13a). Focus on the applied object is implied in the description of (13a); nothing is said about focus in the caustive (13b). Although relevant data are lacking, it's implied that the rule designed for (12) cannot be collapsed with “Perfective Tone Loss” i.e. that even with an object clitic an unauxiliated recent past still lacks stem H, as in hypothetical (i) based on Odden (1996, 74, ex 112):

   (i) \[ C_{LS} \; telek-\; i \quad \text{Manomboono \quad kindiidoolo,} \quad \text{‘... recently cooked sweet potato(es) for M. \quad (= n)’} \]

18. Apart from progressives, no other “noun focal tenses” are mentioned in the source.

19. Cf. Kisseberth & Mmusi (1990) for similar facts in Setswana. The polysynthetic view of applicatives (Baker 1988b, c), deriving the “verb extension” as a lexical head raised from an abstract PP in a downward lexical cascade (Baker 1988a). Byarushengo (2003) reviews problems for applicative incorporation—problems eluded by the inverse architecture in (6b) and (12) where the applicative licensor is above, not below, its spellout position, consistent with upwardly exploded treatments of clausal superstructure (Rizzi 1997, Manzini & Savoia 1998; Cinque 1999).

20. In rightward H displacement forms like (7 b), Goldsmith marks the onsetless aux [a] as “a ‘post-High’ morpheme” (1985, 123).

21. Note the scopal difference between (12) and (13a). In lieu of a phrasing for (13), I offer the slogan flat syntax, flat prosody.

22. The phrase structure of conjunction is notoriously murky, but the two environments potentially form a natural class (Williams 1989). The similarity is underlined by the accentuation of the scrambled object in (14c).

23. “In general, for all verb-focused tenses there is an H-tone on the last mora of the tense-aspect prefix” (Odden 1996, 194).

24. For example if the object in (15a) is anaphoric; Kenesei (2005) argues that narrowing is the only way to get narrow V focus.

25. Odden (1996, 225) gives examples in which descriptive shortening applies twice in one form: once to the root vowel and once to the complex of applicative extension plus tense suffix. However, the likely operation of vocalic ephenthesis in the latter context, though little explored in Bantuist literature, places the theoretical relevance of shortening in doubt. Instrumental data may clarify the correlation between length and pitch dynamism, along the lines of the Chinese literature referred to above. Odden speaks of shortening “providing a second argument for syntactic constituency in potentially unclear cases” (1996, 233), hinting at acoustic unclarity.

26. Examples tweaked. The presence of subject focus in (18c) is not indicated in Cardinaletti’s data but is implied in descriptive remarks (2002, 32). Emarginazione is the closest Italian counterpart in to Germanic “metrical invisibility” (Zubizarreta 1998, 49). Although the typical example in both languages is anaphoric/topical, the requirement is less stringent as shown by (18c) and by (i) from Cinque (1993, 255 fn. 19 citing G. Grewendorf p.c.; cf. also Wagner 2005, 211 and references cited there):

   (i) ... daß der Ärzt bereitwillig \[ FP \quad \text{ jeden Patienten \ untersuch-te} \],
            that the doctor willingly every patient investigated
   ‘... that the doctor willingly examined every patient’

In (18c), nontopical material in XP is deaccented, and in (i) the transitive verb fails to deaccent despite occupying final position.

27. My treatment of Kimutuumbi differs notationally from Pulleyblank’s (1982) use of a Halle & Vergnaud accentual grid plus tonemes. The similarity is that reference to accent lets Pulleyblank avoid rules which first assign tones lexically and then delete them in phrasal contexts. A substantive difference—ignored here—stems from Pulleyblank’s emphasis on nominal prosody.

28. Hyman (1999) abstracts away from automatic vowel elision and glide formation, both of which are marked as apostrophes by Byarushengo & al. (1976) as in (20). Hyman’s glosses for (21c/d) contain the apparent typo “have counted” for ‘are counting’.

29. Right dislocated items were accordingly described as either “afterthoughts” or “recapitulation” (1976, 201), or a broad-focus sentence—like (ii) from Hyman & Byarushengo (1984, 94), contain no instances of [HL], but it’s not clear how the 1999 analysis can avoid generating [HL] there, or on the finite verbs in (20c/d) for that matter, while still correctly producing [HL] contours on the verbs in broad focus sentences like (21 b/c). (Throughout, the verb root belongs to the prelinked H class.)

       K. women CL CL see
       ‘Kakulu, the women (they) see him’

       women K. CL CL see
       ‘The women, Kakulu, see him’

30. In (22), nuclear stress is transcribed as underlining.

31. On the same page, Odden floats a second, more technical solution, without enthusiasm.
References


Russell, K. [1999]. What’s with all these long words anyway? MIT Occasional Papers in Linguistics 17,119-130.


———. [2004]. Conjoint & disjoint; grammaticalized information structure in Bantu. Universiteit van Amsterdam, 3 December.


Sharman, J. & A. Meesussen. [1955]. The representation of structural tones, with special reference to the tonal behavior of the verb, in Bemba, Northern Rhodesia. Africa 25, 393-404.


