Autosegmental representations in an HPSG of Hausa

Berthold Crysmann

Universität Bonn & Universität des Saarlandes

Draft — July 2009

Outline

Hausa

- Chadic sub-branch of Afroasiatic language family
- Over 30 million speakers in Northern Nigeria and bordering areas of Niger
- Tone language
  - Two underlying tones: H, L
  - Three grammatical surface tones: H(igh), L(ow), HL (fall)
  - Rising tone mere result of interaction with intonation (Inkelas and Leben, 1990)
- Both tone and vowel length are distinctive

Tone

- Tone is lexically distinctive
  Minimial pair (Wolff, 1993):

  (1) a. *fàrī* — 'look (n)'
  b. *farȉ* — 'dry season'
  c. *farī* — 'white/whiteness'

- Tone also used to distinguish grammsatical categories, e.g.,
  - TAM

  (2) a. ya zō — he came (relative completive)
  b. yà zō — he should come (subjunctive)

  (3) a. yā zō — he came (absolute completive)
  b. yâ zō — he might come (potential)

Length

- Minimial pair for lexically distinctive vowel length (Newman, 2000)

  (4) a. *fāsà* — 'postpone'

  b. *fasà* — 'smash'

- Vowel length serves to distinguish grammatical notions, e.g. TAM categories

  (5) a. yā zō — he came (absolute completive)

  b. ya zō — he came (relative completive)

Latin script

- Standard orthography (Boko) does not represent tone nor length
- Glottalised consonants distinguished in principle from non-glottalised ones, either graphemically (ɓ, ɗ, ƙ) or by digraphs (ts)
- Tone and length marking restricted to scientific and educational work on Hausa
  - Tone typically marked by diacritics
    - Often, only one tone is marked, the quality of the unmarked tone has to be inferred
    - Both high-marking and low-marking strategies can be found
  - Length either marked by diacritics (e.g., a macron) or double vowels
  - Learner's grammars and course material typically follow scientific practice

Ajami

- Hausa traditionally written using a variant of the Arabic script
- Glottalised consonants (ɓ,ɗ,ƙ,'y,ts) systematically distinguished graphemically from non-glottalised consonants
- Ajami, like the standard Arabic script, marks length distinctions:
  - Long vowels use a combination of letters and vocalisation diacritics
    ya (ي) for front vowels, wau (و) for round back vowels, and alif for /a/ (ا)
  - Short vowels are marked by vocalisation diacritics only
- Ajami distinguishes 5 vowel qualities (Arabic: 3), using additional diacritics for /o/ and /e/
- Tone is not marked

Tone-integrating Suffixes I

- Many inflectional suffixes in Hausa are tone-integrating
  - Segments of the base are preserved
  - Vowel length of the base is generally preserved
  - Tone is holistically assigned to the derived word
- Example: Hausa Plural Class I

(6) -ōXī (H) (Class I)

a. gulà (HL) — gulōlī 'drum stick'

b. tāgà (HL) — tāgōgī 'window'

c. gyàlè (LL) — gyalōlī 'shawl'

d. tàmbayà (LHL) — tambayōyī 'question'

e. kamfànī (HLH) — kamfanōnī 'company'

f. kwàmìtîi (LLHL) — kwamitōcī 'committee'

Tone-integrating Suffixes II

- Example: Hausa Plural Class II

(7) -ai (LH) (Class II)

a. àlhajī̀ (LHL) — àlhà̀zai 'Hadji'

b. ɗālìbī (HLH) — ɗà̀lìbai 'pupil'

c. sankacȅ̄ (HHL) — sànkàtai 'reaped corn laid down in a row'

d. àlmùbazzàrī (LLHLH) — àlmùbàzzàrai 'spendthrift'

- Tone assignment procedes right-to-left, with automatic spreading

Tone-less prefixes I

- Pluractional verbs
- Prefixation of a tone-less reduplicative prefix

(8) C_1VC_1-

a. darnàcē (HLH) — daddarnàcē (HHLH) ‘press down/oppress (gr 1)’

b. karàntā (HLH) — kakkaràntā (HHLH) ‘read (gr 1)’

c. dā̀gurà (LHL) — dàddā̀gurà (LLHL) ‘gnaw at (gr 2)’

d. gyā̀ru (LH) — gyàggyā̀ru (LLH) ‘be well repaired (gr 7)’

Tone-less prefixes II

- Bisyllabic bases:

(9) a. tākà (HL) — tattàkā (HLH) ‘step on (gr 1)’

b. jèfā (LH) — jàjjēfà (LHL) ‘throw at (gr 2)’

Non-integrating Affixes I

- Local segmental and suprasegmental adjustments
- E.g., floating tone-initial *-waa*

(10)
a. karàntā — karàntâwā 'read (gr1)'
b. sayar — sayârwā 'sell (gr5)'
c. kāwō — kāwôwā 'come (gr6)'
d. kāmà — kāmàwā 'catch (gr1)'
e. gyàru — gyàruwā 'be repaired (gr7)'

Non-integrating Affixes II

- Vowel shortening on closed syllables
  - Permissible syllable structures: CV, CVV, CVC
  - Affixation of coda consonants to long vowels triggers shortening

(11) a. ƙwai — ƙwa-n-tà '(her) egg'

b. rɪ̀gā — rɪ̀ga-r-tà '(her) gown'

c. mōtà̄ — mōtà-r-tà '(her) car'

(12) a. ƙwai — ƙwâ-n 'the (aforementioned) egg'

b. rɪ̀gā — rɪ̀gâ-r 'the (aforementioned) gown'

c. mōtà̄ — mōtà-r '(her) car'

Non-integrating Affixes III

(13) Fàransà 'France' — Bàfaranshè (m), Bàfaranshìyā (f) , Faransāwā (pl) 'French'

(14) Jāmùs 'Germany' — Bàjāmushè (m), Bàjāmushìyā (f) , Jāmusāwā (pl) 'French'

(15) a. Pàlàsɗīnù 'Palestine' — ɗan/mùtumìn Pàlàsɗīnù (m) — Palasɗīnāwā (pl) 'Palestinian'

b. Bosniyà 'Bosnia' — ɗan/mùtumìn Bosniyà (m) — Bosniyāwā (pl) 'Bosnian'

Orthographemics in the LKB I

- String unification approach to orthographemics lends itself naturally to
  - Gemination
  - Consonant reduplication
  - Partial reduplication

(16)
```
%(wild-card (?v aeiou))
%(letter-set (!c bcdfghjklmnpqrstvwxyzɓɗƙ́))
noun_pl1_vow_ir :=
    %suffix (!c?v !co!ci) (t?v toci)
    (s?v soshi) (w?v woyi) (ts?v tsotsi)
    noun-plural-infl-rule &
    ...
```

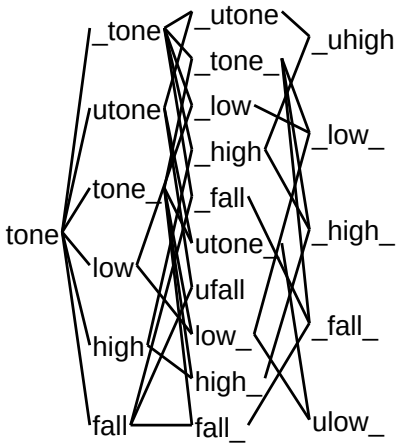
- Elegant treatment of segmental morphophonology necessitates separation of segmental and tonal information

Orthographemics in the LKB II

- Separation of segmental and suprasegmental information performed by a preprocessing step
- RE preprocessor converts diacritics into suffixal tone/length annotations (in reverse order)
- Suffixal annotations subsequently converted into feature structure representations using LKB orthographemic rules
- Explicit tone marking is preserved (suffix _H and _L), implicit marking is represented by underspecified _*
- Explicit length marking is marked by colon
- Examples:
  - ɗàlìbai ↦ ɗalibai_*_L_L:
  - ɗalibai ↦ ɗalibai_*_*_*
  - ɗaalibai ↦ ɗalibai_*_*_*:

- Marking strategy needs to be inferred from the input (high-marked, low-marked, unmarked etc.)
- Suprasegmental interpretation of unmarked items depends on the marking strategy used for marked items (if any)
- Type hierarchy recognises linguistic tone (*high, low, fall*), as well as marking regimes tone (*tone_* vs. *_tone*)
- Overt marking (e.g. tone_ for high-marking regime) restricts interpretation of unmarked tone

(17)



- Orthographemic rules convert suffixal tonal annotations into feature structure representations
- Overt marking of tone or length restricts the marking regime

(18)

```
_HH_ir :=
  %suffix (* _H:)
  diacritic-irule &
  [SUPRA [TONE [LIST #tones,
                LAST #tl],
          LEN  [LIST #lens,
                LAST #ll]],
   DTR [SUPRA [TONE [LIST
                high-marked-list &
                <high . #tones>,
                LAST #last],
              LEN  [LIST
                long-marked-list &
                <long . #lens>,
                LAST #ll]]]].
```

```
_*_ir :=
  %suffix (* \*)
  diacritic-irule &
  [SUPRA [TONE [LIST #tones,
                LAST #tl],
          LEN  [LIST #lens,
                LAST #ll]],
   DTR [SUPRA [TONE [LIST
                <utone . #tones>,
                LAST #last],
              LEN  [LIST
                <ulength . #lens>,
                LAST #ll]]]].
```

- List constraint (`high-marked-list`) propagate marking regime across word's TONE list
- Restriction to `tone_` disambiguates unmarked tone (`utone`) to `low`

(19)

```
high-marked-list :=
  tone-marked-list.
high-marked-null :=
  high-marked-list &
  tone-marked-null.
high-marked-cons :=
  high-marked-list &
  tone-marked-cons &
  [FIRST tone_,
   REST high-marked-list].
```

```
hm-llist := tm-llist.
hm-clist := tm-clist &
    hm-llist &
    [FIRST high-marked-list,
     REST hm-llist].
hm-nlist := hm-llist & tm-nlist.
```

- `hm-llist` propagates marking regime across sentence (list of lists) (left to right (_LTONE) and right to left (_RTONE))

(20)
a. Fully unspecified: `Ya zo` (4 readings: *yā zō*, *ya zō*, *yà zō*, *yâ̂ zō*)
b. Length specified: `Ya zoo` (2 readings: *ya zō*, *yà zō*)
c. Length specified: `Yaa zoo` (1 reading: yā zō)
d. Tone/length specified: `Ya kaawoo shì` (1 reading: *ya kāwō shì*)
e. Fully specified: `Yá zóó` (1 reading: *ya zō*)
f. Inconsistent: `Yaa zo` (0 readings)

```
h*-list := list.
h*-cons := h*-list &
           cons & [FIRST high,
   REST h*-list].

h*-null := h*-list & null.

h*-l-list := list.
h*-l-cons := h*-l-list &
             cons & [FIRST low,
                     REST h*-list].
```

(21)

```
noun_pl1_vow_ir :=
%suffix (!c?v !co!ci) ...
noun-plural-infl-rule  \&
[SUPRA
   [TONE [LIST h*-list],
    LEN [LIST < long, long . #ll>,
         LAST #llast] ],
 DTR [SYNSEM.LKEYS.--MCLASS n-pl-1,
      SUPRA.LEN [LIST  < [] . #ll>,
                 LAST #llast]]].
```

(22)

(23)

```
f-sg-noun_def_high_ir :=
 %suffix (!v !vr) (!vi !vr) ...
 noun-def-f-sg-irule &
 [SUPRA [TONE [LIST <fall . #tl >,
               LAST #tlast],
         LEN [LIST <short . #ll>,
              LAST #llast]],
  DTR [SUPRA
        [TONE [LIST <high . #tl>,
               LAST #tlast],
         LEN [LIST <[] . #ll>,
              LAST #llast] ]]].
```

References I

Inkelas, Sharon and William R. Leben. 1990. Where phonology and phonetics intersect: The case of Hausa intonation. In M. E. Beckman and J. Kingston, eds., *Between the Grammar and the Physics of Speech*, Papers in Laboratory Phonology, pages 17–34. New York: Cambridge University Press.

Newman, Paul. 2000. *The Hausa Language. An Encyclopedic Reference Grammar*. New Haven, CT: Yale University Press.

Wolff, Ekkehard. 1993. *Referenzgrammatik des Hausa*. Münster: LIT.