Punctuation is prosody
Making historic transcriptions of Karuk accessible for revitalization and research

Clare Sandy and Line Mikkelsen
University of California, Berkeley

ICLDC 4, Honolulu, Hawai‘i
28 February 2015
Outline

Intro
The Karuk language
Obtaining prosody from archival materials?
Specific research questions

Results
Intonation in Karuk
William Bright transcriptions
J. P. Harrington transcriptions
Correlation

Conclusions
Summary of findings
Further questions
Applications of findings
The Karuk language

- Classification: isolate within alleged Hokan group
- Geography: traditionally spoken along the middle Klamath River (Northern California, US)
- Neighbors
  - Yurok (Algic)
  - Shasta and Chimariko (also Hokan)
  - Tolowa and Hupa (Athabaskan)
- Vitality
  - 1850: estimated 1800-2700 speakers
  - 1950: estimated 100 speakers (Bright 1957)
  - 2015: <10 first-language speakers, 20-50 learners and L2 users
Past documentation

Early 20th century: Kroeber, Harrington, de Angulo & Freeland

- J. P. Harrington materials include:
  - Published texts (Harrington, 1930, 1932a,b)
  - Several reels of microfilm of field notes archived at the Smithsonian
  - One surviving recording of texts spoken by Phoebe Maddux archived at Phoebe Hearst Museum and Library of Congress

More recent: William Bright (1940s–early 2000s), Monica Macaulay, current authors and others.

- William Bright materials include:
  - Published grammar *The Karok Language* in 1957
  - Recordings and fieldnotes archived at Survey of California and Other Indian Languages and American Philosophical Society
General problem

Utterance-level prosody is important ...

- For language teaching and learning
- For understanding syntax-phonology interactions, in particular, word order variation
- For understanding variation in word-level accentuation

... but is difficult to recover from archival materials without audio

- Prosody is often not explicitly transcribed
- Different field workers use various informal methods to represent prosody

Can we interpret archival transcriptions to obtain linguistically useful data on the prosody of utterances?
How do we transcribe prosody? Punctuation, of course!

For example:

There’s Elizabeth.

Where’s Elizabeth?

Elizabeth is where?!

It’s Elizabeth!

As with orthography, different languages use the same symbols to represent different sounds.

Our knowledge of our language’s prosody supplies acceptable phonetic realizations of intonation associated with this shorthand.
Specific Questions

▶ Can we extract useful utterance-level prosodic data from archival Karuk materials?
▶ To what degree can the Bright and Harrington transcriptions be correlated?

Methodology

▶ Historic recordings and transcriptions
Karuk intonation basics

- At most one accent per word
- Intonational phrase (IP) characterized by:
  - gradual rise to peak, followed by fall
  - pitch peak coincides with (typically last) word accent
- End of utterance characterized by:
  - devoicing, glottalization, or accentual effects

Nettie Reuben, 1949, 'Coyote’s Homecoming’ (WB_KL-02a:09)
Bright system

Bright explicitly uses punctuation to indicate intonational contours,

“...associated not with any grammatical categories, but rather with the emotions of the speaker. (Bright 1957:15)”

Punctuation represents the contour of the preceding sequence:

- **period** “associated with lack of marked emotion (Bright 1957:16)”
  rising-falling pitch contour, utterance-final boundary, low boundary tone, pause

- **comma** “associated with lack of marked emotion (Bright 1957:16)”
  rising-falling pitch contour, accentual phrase boundary, weak low boundary tone, optional pause

- **exclamation point** “associated with excitement (Bright 1957:16)”
  higher register for utterance

- **semi-colon** “associated with sadness (Bright 1957:16)”
  falling intonation
Translation into ToBI

Fairly straightforward translation into (a simplified) ToBI (Beckman et al., 2005), for example:

- **Word accents**: $H^*+L$
- **Bright’s “comma pause”**: $wL\%$ for weak L boundary tone, normally corresponds to a Break index 2
- **Bright’s “period pause”**: $L\%$ for L boundary tone, normally corresponds to a Break index 3 with finality (Venditti, 1997) or a Break index 4 (Beckman and Ayers, 1997)
J. P. Harrington audio

Phoebe Maddux, 1929, 'Bluejay Myth' (JPH_PHM-24-343a:03)

- One recording (made by Boas, transcribed by Harrington)
- Intonation seems comparable to that described by Bright
  - Typical pitch rise to a peak, followed by an abrupt fall, and then drops more or less gradually to low
  - Variable based on many factors, e.g., word accentuation, context of utterance, speaker intention
"\text{ukni}^\text{.}
\text{atah}^\text{.} \text{hā riva}
\text{kunvāra}^\text{.} \text{rahitihanik}.
\text{te}^\text{.} \text{pā}^\text{.} \text{nay pay}
\text{te}^\text{.} \text{imay pay}
\text{yī 00}
\text{ukūhe}^\text{.}^\text{e}.
\text{Then after a while}
\text{all at once}
\text{one he got sick.}

\text{1. Bluejay Myth.}
\text{"ukni}^\text{.}
\text{ata hā riva}
\text{kunvāra}^\text{.} \text{rahitihanik.}
\text{\text{ukni}^\text{.} \text{infrared}} \text{thus they were living.}
\text{te}^\text{.} \text{pā}^\text{.} \text{nay pay}
\text{te}^\text{.} \text{imay pay}
\text{yī 00}
\text{ukūhe}^\text{.}^\text{e}.
\text{Then after a while}
\text{all at once}
\text{one he got sick.}
J. P. Harrington transcription – preview

How does his punctuation reflect prosody in that recording?
- Sentences (marked with period) reflect utterances
- In some cases, the Karuk is split into more sentences than the English free translation
- Quotes introduced with colons (but some quoted speech does not follow verb of speaking)
- Occasional commas and semi-colons used inconsistently; a few mismatches between notes and typescript

How much prosody is recoverable from his transcription?
- Utterances are recoverable (using transcription, not free translation)
- Detailed variation among intonational contours not recoverable
Transcription of utterances

Phoebe Maddux, 1929, 'Bluejay Myth' (JPH_PHM-24-343a:9)

1. Period (follows “basic” sentence)
   ▶ Transcribed consistently in Karuk lines\(^1\)
   ▶ Intonation: typically rise to peak, late low fall, devoicing/glottalization at right edge
   ▶ Pause: significant, obvious (419 – 2,220 ms, median 1,434 ms)

---
\(^1\)In notes, period occasionally missing in Karuk transcription but present in English gloss and typed version; line break present.
Transcription of utterances

Phoebe Maddux, 1929, 'Bluejay Myth' (JPH_PHM-24-343a:32-33)

Karuk transcription:  
kúkuum vúra úpkuupha. kári xás kúkuum kunpíkaar.

Original gloss:  
‘They did that way again, went to get her again.’

Revised gloss:  
‘Then she did the same way again. They went to get her again.’

Autonomy of punctuation in free translation indicates that these are units in the Karuk
Transcription of utterances

Phoebe Maddux, 1929, 'Bluejay Myth' (JPH_PHM-24-343a:44-45)

2. Formulaic sentences (i.e., introducing and ending stories)
   ▶ *uknį* ‘once upon a time’ unusual, as expected
   ▶ Others – intonation same as typical sentences; pause can be abbreviated
Use of commas

3. Sentence-internal commas

- Few examples. Always transcribed on one line.
- Intonation: low but no devoicing or glottalization before comma; low afterwards
- Pauses: very short (< 100 ms) or nonexistent (coincident with stop closure released into following word)
- Comma vs. pre- or post-posed material w/o comma?

Phoebe Maddux, 1929, 'Bluejay Myth' (JPH_PHM-24-343a:23)
Use of commas

Phoebe Maddux, 1929, 'Bluejay Myth' (JPH.PHM-24-323a:4-5)

Original transcription:  
uum káru káan úkrii kachakâachich; êem káru apurúvaan.

Revised transcription:  
uum káru káan úkrii kachakâachich. êem káru apurúvaan.

English gloss:  
‘Bluejay was also living there, she was a doctor and sorcerer.’

4. Variably transcribed sentences

- Phrases (notes) → sentences (typescript)
- Intonation: one unusual, others follow basic pattern
- Pauses: fall in range of sentence-level pauses (627 – 1,820 ms, median 1,490 ms)
# Correlating the transcription systems

<table>
<thead>
<tr>
<th>Bright</th>
<th>Harrington</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Utterance</strong></td>
<td></td>
</tr>
<tr>
<td>Period (rise to fall)</td>
<td>Period includes all these</td>
</tr>
<tr>
<td>Semi-colon (falling)</td>
<td></td>
</tr>
<tr>
<td>Exclamation point (high)</td>
<td></td>
</tr>
<tr>
<td><strong>Smaller Intonational Phrase</strong></td>
<td></td>
</tr>
<tr>
<td>Comma</td>
<td>Comma or not marked</td>
</tr>
<tr>
<td><strong>Introducing quote</strong></td>
<td></td>
</tr>
<tr>
<td>Comma (some finality)</td>
<td>Colon</td>
</tr>
<tr>
<td>Not marked (no finality)</td>
<td>Colon</td>
</tr>
<tr>
<td><strong>Quoted speech</strong></td>
<td></td>
</tr>
<tr>
<td>Quotes</td>
<td>Quotes</td>
</tr>
</tbody>
</table>
Summary of findings

Can we extract useful information about intonation from transcriptions? Can they be correlated?

▸ Yes! (although can’t recover everything)
▸ Same intonational contour is most common in recordings from both eras
▸ Where they differ raises interesting questions
Open questions about intonation and accentuation

- Bright’s ‘expressive’ contours also found in Harrington, but:
  - Somewhat different contexts from Bright’s description.
  - Interesting directions to explore re: information structure.
- Question of individual variation/style.
  - Higher register in quoted speech found with one speaker, not with another
  - Individual style? Generational difference?
- Lack of consistency in actual phonetic contours of Bright’s transcriptions:
  - Early pitch peaks in utterances, esp. on circumflex accent – worth looking at again for subtle accentual distinctions
  - Phonetic correlate of ‘comma’ prosody less consistent than ‘period’ – sparsity of data or less than clear-cut category?
Pedagogy

Karuk utterance prosody is different from English

- Must be learned
- Pedagogical challenge - how to ‘read’ punctuation in a different way
- We can identify a typical utterance contour which can be the baseline for learning/teaching
Syntactic analysis

- Ability to recover intonational phrasing from archival record allows empirical testing of hypotheses about determinants of word order
  - Obligatory post-verbal positioning of complement clauses is prosodically governed
  - Extraposition of non-initial conjuncts is prosodically governed
  - Two distinct post-verbal positions for DPs
Yôotva!

Present and past Karuk elders, teachers, and activists: Tamara Alexander, Lulu Alexander, Sonny Davis, Susan Gehr, Julian Lang, Crystal Richardson, Nancy Richardson, Bud Smith, Vina Smith, Florrine Super, Arch Super; and †Lucille Albers and †Charlie Thom, Sr.

Erik Maier and the rest of the Karuk Study Group at UC Berkeley

This material is based upon work supported by a Mellon Project Grant from the UC Berkeley Division of Arts and Humanities and a National Science Foundation Doctoral Dissertation Research Improvement Grant under Grant Number (BCS-1349075)
References


Phoebe Maddux, 1929, 'Bluejay Myth' (JPH_PHM-24-323a:35)

5. Colon (follows verb of speaking, before quoted speech)
   - Pauses: short (67 – 741 ms, median 698 ms)
   - Intonation: flat or slight drop, not to extreme L
   - Changed mind about whether songs ‘count’ as quoted speech
   - Compare: Bright uses either no punctuation or a comma
Punctuation in quoted contexts

6. Quoted speech
   - Following verb of speaking: double quotes in typescript glosses, not in free translation or notes
     
     **Gloss**  Then again indeed thus she said: “Her indeed somebodies they are thinking it.”
     **Free**  Then she said: Somebody is making it.

   - Without overt verb of speaking: double quotes in typescript free translation, not in glosses or notes
     
     **Gloss**  Then now it hazel nuts give me.
     **Free**  “Give me hazel nuts this time.”

   - Pauses: well within range of “basic” sentences
   - Intonation: peaks indicate info structure, not raised overall

7. Lines of song
   - Punctuation: variable (notes); commas (typescript)
   - Pauses: shorter (334 – 876 ms, median 517 ms)
   - Intonation within sung context not analyzed