Covariance of syntactic and phonological contact effects in Eastern Cham

Kenneth Baclawski Jr.
University of California, Berkeley
NWAV 45, November 6, 2016
Intro

• Two phonological variables covary with one syntactic variable
  • Categorical presence or absence of construction
  • In contrast with alternations typical of socio-syntax
• Eastern Cham contact with Vietnamese has phonological and syntactic exponents
• Further evidence for the study of syntactic variation (Grondelaers & Speelman 2007, a.o.)
• Generally supports a competing grammar approach (cf. Kroch 1994)
Outline

• Eastern Cham/Vietnamese bilingualism
• Phonological variables
  • /r/, /j/
  • /ŋ/
• Syntactic variation
  • Discourse anaphora
  • Wh-topicalization
• Covariance and socio-syntax
Eastern Cham & Vietnamese

- Eastern Cham
  - Austronesian: South-Central Vietnam
  - 100,000 speakers, all bilingual with Vietnamese
  - Endangered (UNESCO)
    (cf. Brunelle & Vǎn Hǎn 2015)

- Vietnamese
  - Austroasiatic
  - 100 million speakers worldwide
  - Dominant political/socioeconomic language of Vietnam

- Typological similarities due to thousands of years of contact (Thurgood 1999)
Eastern Cham diglossia

• Exhibits stable diglossia
  • Brunelle (2009) on one variable: monosyllabification
  • Baclawski (2016) reports two other phonological variables

• Register #1: Formal speech with connections to Cham script

• Register #2: Colloquial speech with Vietnamese contact effects
Predictions

• Intense language contact can lead to morphosyntactic contact effects (e.g. Thomason & Kaufman 1988)

• In a unidirectional contact situation, contact effects are liable to have parallel distributions
Data collection

- 30 native Eastern Cham speakers
  - 15 male, 15 female
  - Aged 18-79 (median = 22)
  - Raised in Cham villages, Ninh Thuận Province
- Survey targeting phonological variables
- Syntactic elicitation
Phonological variables

• 1. /r/, /j/
• 2. /ŋ/ / V_{rd} –
• Both are Vietnamese contact effects (Baclawski 2016)
• Both exhibit significant inter- and intra-speaker variation
/r/, /j/ in Eastern Cham

• Blood (1961) reports variation of /r/, /j/ in onsets:
  • [r] ~ [j] ~ [z]

• Baclawski (2016) reports similar facts:
  • [r] ~ [rʒ] ~ [j] ~ [z] ~ [jʒ]

• /rːàʔ/ ‘market’
  • [rːàʔ] ~ [rʒːàʔ] ~ [jːàʔ] ~ [zːàʔ] ~ [jʒːàʔ]

• /jːaw/ ‘wood’
  • [rːaw] ~ [rʒːaw] ~ [jːaw] ~ [zːaw] ~ [jʒːaw]
/j/ in Vietnamese

• Tran & Norris (2010) report for /j/ in onsets:
  • [j] (Southern dialects) ~ [z] (Northern dialects) ~ [ʐ] (observed in Central)
  • Baclawski (2016) conjectures that all these forms are present in South-Central Vietnam
    • Dialectology is urgently needed in Vietnam in general

• /jèː/ <giò> ‘hour’: [jèː] ~ [zèː] ~ [ʐèː]
/ŋ/ / V_{rd~} in Eastern Cham

- Blood (1961) reports variation of /ŋ/ / V_{rd~}:
  - [ŋ] ~ [ŋm]

- Baclawski (2016) reports the same facts

- /pùŋ/ ‘top.of’
  - [pùŋ] ~ [pùŋm]
/ŋ/ / V_{rd} – in Vietnamese

• Allophony of /k,ŋ/ reported since the 1600’s (Jacques 2002, citing the Portuguese missionary de Rhodes; Thompson 1965)
  • /k,ŋ/ → [kp, ŋm] / V_{rd} –

• /ɔ́k/ <óc> ‘snail’ → [ɔ́kp]
• /oŋ/ <ông> ‘grandfather’ → [oŋm]
  • Note: Eastern Cham lacks final /k/

11/6/2016
Baclawski Jr., NWAV 45
12
Summary

• For /r/, /j/: [j], [z], and [jʒ] are contact variants
  • [r] and [rʒ] are non-contact variants

• For /ŋ/: [ŋm] is a contact variant
  • [ŋ] is a non-contact variant

• It seems clear that this is due to language contact in the sense of Thomason (2008)

• Further research is needed to establish this without a shadow of a doubt (cf. Poplack & Levey 2010)
Survey

- Conducted in Ho Chi Minh City, 2015
  - Word list (n=50)
  - Sentences (n=50)
  - Consultants asked to speak as they do at home
    → Colloquial register
Results: Gradience

- Intra-speaker variation
- Speaker MXL:
  - [jːàʔ] ‘market’ (Word List)
  - [rːàʔ] ‘market’ (Sentence)

<table>
<thead>
<tr>
<th></th>
<th>Non-contact</th>
<th>Both contact/ non-contact</th>
<th>Only contact</th>
</tr>
</thead>
<tbody>
<tr>
<td>/r/,/j/</td>
<td>2</td>
<td>21</td>
<td>7</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Non-contact</th>
<th>Both contact/ non-contact</th>
<th>Only contact</th>
</tr>
</thead>
<tbody>
<tr>
<td>/ŋ/ / V_{rd}</td>
<td>Ø</td>
<td>28</td>
<td>2</td>
</tr>
</tbody>
</table>
Results: Gradience

• Variation is not obviously lexical diffusion

<table>
<thead>
<tr>
<th>Word</th>
<th>Non-contact</th>
<th>Contact</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>/rːàʔ/ ‘market’</td>
<td>20</td>
<td>16</td>
<td>36</td>
</tr>
<tr>
<td>/kra/ ‘monkey’</td>
<td>28</td>
<td>25</td>
<td>53</td>
</tr>
<tr>
<td>/krɔŋ/ ‘river’</td>
<td>28</td>
<td>20</td>
<td>48</td>
</tr>
<tr>
<td>/hrɛj/ ‘today’</td>
<td>16</td>
<td>21</td>
<td>37</td>
</tr>
<tr>
<td>/kyɔw/ ‘wood’</td>
<td>19</td>
<td>25</td>
<td>44</td>
</tr>
<tr>
<td>/pːrɔj/ ‘yesterday’</td>
<td>26</td>
<td>29</td>
<td>55</td>
</tr>
</tbody>
</table>
Results: Gradience

• Variation is not obviously lexical diffusion

<table>
<thead>
<tr>
<th>Word</th>
<th>Non-contact</th>
<th>Contact</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>/ɲɔ̀ŋ/ ‘circle’</td>
<td>24</td>
<td>3</td>
<td>27</td>
</tr>
<tr>
<td>/ʔɪŋ ʔɔ̀ŋ/ ‘frog’</td>
<td>36</td>
<td>18</td>
<td>54</td>
</tr>
<tr>
<td>/θɔ̀ŋ/ ‘knife’</td>
<td>6</td>
<td>51</td>
<td>57</td>
</tr>
<tr>
<td>/kɾɔŋ/ ‘river’</td>
<td>24</td>
<td>35</td>
<td>59</td>
</tr>
<tr>
<td>/pùŋ/ ‘top.of’</td>
<td>28</td>
<td>28</td>
<td>56</td>
</tr>
<tr>
<td>/thɔŋ/ ‘with’</td>
<td>5</td>
<td>38</td>
<td>43</td>
</tr>
</tbody>
</table>
Results: Gradience

- Not obviously a lexical implicational hierarchy
- If a speaker only uses a variable in one lexical item:

<table>
<thead>
<tr>
<th></th>
<th>[r]</th>
<th>[j]/[z]</th>
<th>[ŋ]</th>
<th>[ŋm]</th>
</tr>
</thead>
<tbody>
<tr>
<td>/krɔŋ/</td>
<td>‘river’</td>
<td>/kyɔw/ ‘wood’</td>
<td>/vɔŋ/ ‘circle’</td>
<td>/θɔŋ/ ‘knife’</td>
</tr>
<tr>
<td>/pːrɔj/</td>
<td>‘yesterday’</td>
<td>/kyɔw/ ‘wood’</td>
<td>/ʔɪŋ ʔɔŋ/ ‘frog’</td>
<td>/θɔŋ/ ‘with’</td>
</tr>
<tr>
<td>/pːrɔj/</td>
<td>‘yesterday’</td>
<td>/pùŋ/ ‘top.of’</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>/θɔŋ/ ‘with’</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Results: Village

- Likelihood ratio tests for /r/ and /ŋ/ (cf. Baayen 2008)
  - Random effects: Speaker, Lexical item
  - Factors: Age, Gender, Task (e.g. Word List), Village

- /r/: Village significant (p < 0.05)
  - Age, Gender, Task n.s.
- /ŋ/: Village significant (p < 0.05)
  - Age, Gender, Task n.s.
Results: Village

• Two (of 8) villages predicted [j]/[z] & [ŋ̃m]: Palei Hamu Tanran & Palei Hamu Craok

• Baclawski (2016) claims that these two villages have greater socioeconomic contact with Vietnamese communities
Interim summary

• Two phonological contact variables

• Gradient inter- and intra-speaker variation

• Significance of speaker village, but much variation left to be explained
Syntactic variable

- Optional *wh*-topicalization
  - Licit for some Eastern Cham speakers
  - Ungrammatical in Vietnamese

(1) \{\%\} zut ?ːa \{thɛj\} băŋ lːɔ mːɔ
   friend invite who eat meat cow
   ‘Who did you [friend] invite to eat beef?’

(1’) \{\*\} Bạn mòi \{ai\} ăn thịt bò?
   friend invite who eat meat cow
   ‘Who did you [friend] invite to eat beef?’
Information structure & syntax

• Topic- (and focus-)movement is seen as the result of prosodic factors, or features in the lexicon

• Optional ‘p-movement’/prosodic movement

• [topic] features in the syntax
  • Rizzi (1997)’s Cartographic approach, López (2009), etc.
  • Optionally assigned (Mikkelsen 2005; cf. Bolinger 1972)
Information structure & syntax

- Topic feature in the syntax
- Prosodic constraint/rule: e.g. topics must be in a separate intonational phrase
Topic as discourse anaphora

- There are many notions of ‘topic’
  - A. Aboutness
  - B. Old information (cf. Reinhart 1981)
  - C. Discourse relevant

- López (2009) discards A,B
  - Supports a ‘discourse subordination test’
    (Asher & Lascarides 2002; Asher & Vieu 2005)
Topicalization

• Discourse coordination context (‘X, then Y’):
  • Catalan: × clitic right-dislocation
  • Eastern Cham: × topicalization

Context: ‘He cooked the meat.’

(2) #Després se la va menjar, la carn.
afterwards CL CL.ACC PAST eat.PART the meat
‘Afterwards, he ate the meat.’ (López 2009: 48)

Context: ‘My older brother cooked chicken.’

(2') #plɔh năn, lɔ nũʔ, aj băŋ
after.that meat chicken o.s. eat
‘After that, my older brother ate the chicken.’
Topicalization

• Discourse subordination (‘X, because Y’):
  • Catalan: ✓ clitic right-dislocation
  • Eastern Cham: ✓ topicalization

CONTEXT: ‘Look at him cooking the meat.’
(3) La fà molt be, el Joan, la carn.
   CL.ACC make very well the Joan the meat
   ‘He [Joan] cooks the meat very well.’ (López 2009: 49)   Catalán

CONTEXT: ‘Look at him cutting the meat.’
(3’) rālo, nu sîʔ ffjem lo
   meat 3SG chop good very
   ‘He cuts the meat very well.’   Eastern Cham
Wh-topicalization

- The discourse subordination test also applies to (D-linked) wh-phrases in Eastern Cham
- cf. Pan (2016) on Mandarin

_CONTEXT: A: ‘I ate dinner.’

(4) #hwăʔ ph̥lɔh, ֿkeit, hi ḷǎʔ?
    eat.rice after what 2SG do
    B: ‘After eating, what did you do?’ [Coordination with A]

(4′) keit, zut băn̥ŋ
    what friend eat
    B: ‘What [kind of food] did you eat?’ [Subordination with A]
Topicalization in Vietnamese

- Topicalization is marked overtly with thi
- Wh-topicalization seems to be impossible

Context: A: ‘I ate dinner.’

(4) #Sau khi ăn tối, (cáí) gì (thì) bạn làm?
  after eat.dinner CLF what TOP friend do
  B: ‘After eating dinner, what did you do?’ [Coordination with A]

(4’) #(Cáí) gì (thì) bạn ăn?
  CLF what TOP friend eat
  B: ‘What [kind of food] did you eat?’ [Subordination with A]
Syntax/discourse elicitation

- Speaker information (7 speakers)
- Elicitation of 4 discourse coordination, 4 discourse subordination contexts

<table>
<thead>
<tr>
<th>Context</th>
<th>Discourse coordination</th>
<th>Discourse subordination</th>
</tr>
</thead>
<tbody>
<tr>
<td>A: ‘First, I cooked dinner.’</td>
<td>B: ‘Then, <strong>what</strong> did you do?’</td>
<td>B: ‘<strong>What</strong> did you cook?’</td>
</tr>
<tr>
<td>A: ‘First, I invited someone.’</td>
<td>B: ‘Then, <strong>who</strong> did Kenny invite?’</td>
<td>B: ‘<strong>Who</strong> did you invite?’</td>
</tr>
<tr>
<td>A: ‘I ate dinner.’</td>
<td>B: ‘<strong>What</strong> did you eat?’</td>
<td>B: ‘<strong>What</strong> did you eat?’ [Clarification]</td>
</tr>
<tr>
<td></td>
<td>[Elaboration]</td>
<td>B: ‘Then, <strong>what</strong> is she eating?’</td>
</tr>
<tr>
<td>A: ‘She isn’t eating meat.’</td>
<td>B: ‘<strong>What</strong> does she like to eat?’</td>
<td></td>
</tr>
</tbody>
</table>
Variation in grammaticality

• 5 speakers accepted *wh*-topicalization
• 2 speakers categorically rejected it

**Context:** A: ‘First, I invited someone [to come eat beef with us].’

(5) *plɔh, thej, zut ʔ:a* [Subordination with A]
before who friend invite
B: ‘Who did you [friend] invite, then?’ (Speaker NNA)

(5') *#thej, zut ʔ:a bəŋ lːo mːo* [Subordination with A]
who friend invite eat meat cow
B: ‘Who did you [friend] invite to eat beef?’ (Speaker NTNT)
Overgeneralization

• 2 speakers accepted *wh*-topicalization in all contexts (but did not necessarily produce it)

**CONTEXT:** A: ‘I just ate dinner.’

(5) ḵeit, jut ḵja hwāʔ? [Coordination with A]
what friend just eat.rice

B: ‘What did you [friend] eat? [Clarification]’ (Speaker TQD)

**CONTEXT:** A: ‘That woman is cooking duck.’

(6) ki ḵo nǐʔ m:ej năn ṭɔʔ n̥ãʔ? [Coordination with A]
what REL CLF woman that PROG make[cook]

B: ‘What is that woman cooking? [Clarification]’ (Speaker DPNS)
Results: Syntactic variation

- Varying acceptability of \textit{wh}-topicalization

<table>
<thead>
<tr>
<th></th>
<th>Group 1a</th>
<th>Group 1b</th>
<th>Group 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>\textit{wh}-topicalization</td>
<td>✓  ✓  ✓  ✓*  ✓*  ✗  ✗</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Contact with VN?</td>
<td>✗  ✗  ✗  ?  ?  ✓  ✓</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Covariance

- There is a weak covariance between the phonological and syntactic variables

<table>
<thead>
<tr>
<th></th>
<th>Group 1a</th>
<th>Group 1b</th>
<th>Group 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NNA</td>
<td>ER</td>
<td>TDK</td>
</tr>
<tr>
<td>/r/,/j/ → [j]/[z]</td>
<td>18%</td>
<td>36%</td>
<td>100%</td>
</tr>
<tr>
<td>/ŋ/ → [ŋm]</td>
<td>35%</td>
<td>47%</td>
<td>5%</td>
</tr>
<tr>
<td>wh-topicalization</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>
Covariance

- There is a weak covariance between the phonological and syntactic variables:
  - If *wh*-topicalization is ungrammatical, speakers predominantly use contact phonological variants.
  - If non-contact phonological variants are predominant, *wh*-topicalization is grammatical.

<table>
<thead>
<tr>
<th></th>
<th>Group 1a</th>
<th>Group 1b</th>
<th>Group 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NNA</td>
<td>ER</td>
<td>TDK</td>
</tr>
<tr>
<td>/r/ → [j]/[z]</td>
<td>18%</td>
<td>36%</td>
<td>100%</td>
</tr>
<tr>
<td>/ŋ/ → [ŋm]</td>
<td>35%</td>
<td>47%</td>
<td>5%</td>
</tr>
<tr>
<td><em>wh</em>-topicalization</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
</tbody>
</table>

*Note: ✔ indicates a high percentage of occurrence, ✔* indicates a moderate percentage of occurrence, and ❌ indicates a low percentage of occurrence.*
Covariance

• There is a weak covariance between the phonological and syntactic variables:
  • Group 2: More contact in syntax $\rightarrow$ more contact in phonology
  • Group 1a: Less contact in phonology $\rightarrow$ less contact in syntax

<table>
<thead>
<tr>
<th>/r/ $\rightarrow$ [j]/[z]</th>
<th>/ŋ/ $\rightarrow$ [ŋm]</th>
<th>wh-topicalization</th>
</tr>
</thead>
<tbody>
<tr>
<td>NNA ER TDK TQD DPNS Average</td>
<td>TTL NTNT</td>
<td></td>
</tr>
<tr>
<td>18% 36% 100% 50% 80% 57%</td>
<td>66% 100%</td>
<td></td>
</tr>
<tr>
<td>35% 47% 5% 30% 60% 59%</td>
<td>100% 74%</td>
<td></td>
</tr>
<tr>
<td>✓ ✓ ✓ ✓* ✓ ✓*</td>
<td>✓ ✓</td>
<td></td>
</tr>
</tbody>
</table>
Covariance & salience

• Salience could be a crucial factor
• Diglossia is very dominant in language attitudes, could mask language contact effects
• These non-diglossic phonological variables are conspicuously non-salient
  • No speakers have offered metalinguistic commentary, even when confronted with the forms (Baclawski 2016)
• The syntactic variable is likewise non-salient
An asymmetry

- *Wh*-topicalization is always optional, possibly obscuring speaker salience
- Group 2 speakers never utter infelicitous utterances to Group 1
- *Wh*-topicalization is infelicitous to Group 2

<table>
<thead>
<tr>
<th>Discourse subordination context</th>
<th>Group 1</th>
<th>Group 2</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>wh</em>-in-situ</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td><em>wh</em>-topicalization</td>
<td>✓</td>
<td>×</td>
</tr>
</tbody>
</table>
Covariance & salience

• Perhaps all these contact effects are due to attitudes toward bilingualism
  (cf. Matras 2009)
  • Group 1a inhibits replication of Vietnamese syntax and, for a subset of speakers, phonology
  • Group 2 does not inhibit replication of Vietnamese syntax or phonology
The continuing search for covariance

• At least for intense language contact situations, syntactic variation can covary with phonological variation

• Much future research is needed:
  • Perception studies to assess salience
  • Larger sample size for discourse elicitation
  • Greater study of Eastern Cham and Vietnamese variation and contact
Acknowledgments

• Thank you very much!

• I would like to acknowledge all the Cham people, my research assistant Hamu Ligaih (Đặng Thanh Quốc Thuận) and his professor, Sakaya (Dr. Trương Văn Môn) for their integral role in translation, finding speakers, and the general organization of the study.

• Additional thanks to my Cham teacher, Mohammad Saleh Thiên. Thanks to Andrew Garrett, Keith Johnson, Line Mikkelsen, Peter Jenks, Terry Kaufman, Jim Matisoff, Justin Davidson, Johanna Nichols, Marc Brunelle, Spencer Lamoureux, Amalia Skilton, Elise Stickles, and Anna Jurgensen; special thanks to my research apprentices, Win Htet Kyaw and Nathan Phillip Cahn; and to my trusty fieldwork assistant, Dylan Calhoun.

• This research was made possible by an Oswalt Endangered Language Grant (Survey of California and Other Indigenous Languages, UC Berkeley) from 2014—2015.

• This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE-1106400. Any opinion, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.