IDENTITY, PERFORMANCE, AND CONSCIOUSNESS: 
THE USE OF LOCALLY SALIENT LINGUISTIC FORMS IN A FORMERLY ISOLATED COMMUNITY

1. LANGUAGE AND IDENTITY IN PETTY HARBOUR, NEWFOUNDLAND

- Van Herk, Childs, and Thorburn (fc) examined language in Petty Harbour, Newfoundland:
  - A post-isolated community 15 km southeast of St. John’s.
  - A traditional isolated fishing community → an urbanizing town connected to the province’s capital.
  - Van Herk et al. explored how “traditional linguistic variables compete with newer linguistic forms” in urbanizing Petty Harbour.

- From sociolinguistic interviews with 24 native residents, Van Herk et al. examine:
  - **Stopping of interdental**
    - e.g., dis tine this thing
  - **Non-standard verbal -s marking**
    - e.g., 1 likes having fun with the kids

- Van Herk et al. find that
  - Young women use the local forms the least (old men the most).
    - The young women still use the local forms, and
    - The local forms are SALIENT markers of local identity

2. PERFORMANCE(?)

- So, despite low(er) rates of use, the younger speakers do make at least symbolic use of those marked local forms.
- I here examine the ideas that:
  - Speakers invoke or PERFORM salient local dialect features to assert localized identities (~ Schilling-Estes 1998).

- A corollary of this sort of interpretation is that the local features are less integrated into the younger speakers’ native grammars than they are in the grammars of the older, local-dialect speakers.
- In other words, “PERFORMANCE” is a non-native or semi-native linguistic strategy.

Research Question:

Can we find evidence of this non- or semi-native performance in the actual speech signal, specifically in its sequential temporal patterns?

3. SEQUENTIAL TEMPORAL PATTERNING OF SPEECH: THE HENDERSON GRAPH

- Henderson, Goldman-Eisler, and Skarbek (1966) explored the SEQUENTIAL TEMPORAL PATTERNS in read speech and spontaneous speech and developed a plotting method that, following Levelt (1989), I call a HENDERSON GRAPH (see figure 1).

- Henderson et al. (1966: 208):
  “each passage of spontaneous speech plotted revealed basically the same structure. … the sequential temporal structure revealed an alternating sequence of periods with different speech/silence rations. Periods in which relatively long pauses and short speech utterances occurred together alternated with periods in which relatively short pauses and long speech periods occurred together. The alternations were apparent as regular changes in the rate of acceleration in series of short straight lines which could be fitted to the over-all slope, relatively steep slopes alternating with relatively shallow slopes.”

- Hesitant phases are interpreted as likely locations of speech planning & processing.
4. This Project

- Using Petty Harbour as a test case, I am attempting to use the Henderson Graph method to assess the realization of sociolinguistic variables.
  - Can we learn anything about sociolinguistic variation from this technique?
  - Does the Henderson Slope correlate with variable realizations? Do other temporal features, like speech rate?
    - Do we see differences in this correlation depending on the salience or performedness of the variable form?

5. Data (Subsample) & Variables

- From Van Herk et al.'s interviews, I’m examining here:
  - 1.6 hours of audio
  - 7 speakers
    - 3 old males & 4 young females

<table>
<thead>
<tr>
<th>Table 1. Speakers in subsample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group</td>
</tr>
<tr>
<td>Speaker*</td>
</tr>
<tr>
<td>Included</td>
</tr>
</tbody>
</table>

* Speaker s, a Petty Harbour local, interviewed speaker z.

- Examining variables:
  - Stopping of (voiced) interdental (ð) e.g., dis for this
- Velar nasal fronting (ing) e.g., talkin’ for talking

- Non-standard verbal -s marking (-s)
  - Too rare!
  - Van Herk et al. obtain 61 -s marked verbs (out of 1,090) in their entire sample (5.6%).
  - I get only 1/15 for old males, 4/36 for young females…

6. METHODS

- Using Praat (Boersma and Weenink 2008) to obtain highly accurate time-alignment (cf. Kendall 2006-2007), I transcribed sections of each interview (lengths of which are indicated in the “Included” row of table 1).
- I added the transcripts to the Sociolinguistic Archive and Analysis Project (SLAAP; http://ncslaap.lib.ncsu.edu/; Kendall 2007, 2008)
- Using SLAAP, I timestamped, extracted, and coded:
  - All instances of (ing) for the transcribed passages.
  - All instances of (ð) with the follow criteria:
    - Only the, this, that, these, those that were impressionistically fricated or stopped (no affricates, no zeros).
    - Did not count occurrences in potentially ambiguous contexts (e.g., at the).
  - No type-token limits.
- I built a Henderson Graph feature in SLAAP (see figure 2).
- Plots Henderson Graphs from time-aligned transcripts (and coded variables) in the archive.
- Generates slope-lines based on a least-squares estimate of best-fit at user-selected utterance transitional points.
- Using this tool, I generated the slope lines (as illustrated in figure 2) for all HENDERSON SEGMENTS, stretches of talk that span at least two utterances and one pause.
  - Through this I capture and record:
    - SLOPE = Best-fit slope
    - ΔSLOPE = Change in slope from previous Henderson Segment
    - SLOPECOMP = Tertiary variable based on the comparison of a given slope and that speaker’s mean Slope (low, norm, high).
  - Also,
    - R² = Correlation coefficient for line
    - DUR = Duration of Henderson segment
    - PAUSEN = Number of pauses within segment
    - PAUSEDUR = Median pause duration within segment
    - ARTRATE = Median articulation rate within segment (σ/second, not including pauses)
    - SPKRATE = Overall speaking rate for segment (total # σ/duration of total segment)
- I extracted the variable data and temporal data from SLAAP.
- Using R (R Development Core Team 2008), I combined the variable data with the slope data (all variable data is time-stamped so each variable code can be associated with the features of its matrix Henderson Segment).
  - I discarded all variables that did not fall within measured Henderson Segments (e.g., occurred in a single utterance – no slope to measure).
- Using Rbrul (Johnson fc), I submitted the data to various logistic regression analyses…
7. RESULTS: Stopping of interdentals (ð)

Table 2. % stopped and Ns for (ð)

<table>
<thead>
<tr>
<th></th>
<th>the</th>
<th>that</th>
<th>this</th>
<th>those</th>
<th>these</th>
<th>totals</th>
<th>Van Herk et al. +Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Old Males</td>
<td>80/155</td>
<td>25/55</td>
<td>5/7</td>
<td>2/3</td>
<td>0</td>
<td>112/220</td>
<td>N = 163</td>
</tr>
<tr>
<td></td>
<td>51.6%</td>
<td>45.4%</td>
<td>71.4%</td>
<td>66.7%</td>
<td>-</td>
<td>50.9%</td>
<td>80.3%</td>
</tr>
<tr>
<td>Young Females</td>
<td>61/155</td>
<td>16/76</td>
<td>6/13</td>
<td>0/1</td>
<td>0/2</td>
<td>83/247</td>
<td>N = 175</td>
</tr>
<tr>
<td></td>
<td>39.4%</td>
<td>21.1%</td>
<td>46.1%</td>
<td>0%</td>
<td>33.6%</td>
<td>33.6%</td>
<td>42.8%</td>
</tr>
<tr>
<td>totals</td>
<td>141/310</td>
<td>51/131</td>
<td>11/20</td>
<td>2/4</td>
<td>0/2</td>
<td>190/467</td>
<td>N = 946</td>
</tr>
<tr>
<td></td>
<td>45.5%</td>
<td>33.6%</td>
<td>50.0%</td>
<td>50.0%</td>
<td>0.5%</td>
<td>40.7%</td>
<td>40.7%</td>
</tr>
</tbody>
</table>

Table 3. Regression analysis for (ð) for all speakers

<table>
<thead>
<tr>
<th>AGE+GENDER</th>
<th>p = 0.0001</th>
<th>logodds* uncentered weight</th>
<th>tokens</th>
<th>stop/stop+fric</th>
</tr>
</thead>
<tbody>
<tr>
<td>old male</td>
<td>0.348</td>
<td>0.59</td>
<td>220</td>
<td>0.509</td>
</tr>
<tr>
<td>young female</td>
<td>-0.348</td>
<td>0.42</td>
<td>245</td>
<td>0.339</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>WORD</th>
<th>p = 0.0346</th>
<th>logodds uncentered weight</th>
<th>tokens</th>
<th>stop/stop+fric</th>
</tr>
</thead>
<tbody>
<tr>
<td>this</td>
<td>0.518</td>
<td>0.65</td>
<td>20</td>
<td>0.550</td>
</tr>
<tr>
<td>those</td>
<td>0.024</td>
<td>0.54</td>
<td>4</td>
<td>0.500</td>
</tr>
<tr>
<td>the</td>
<td>0.014</td>
<td>0.53</td>
<td>310</td>
<td>0.455</td>
</tr>
<tr>
<td>that</td>
<td>-0.348</td>
<td>0.40</td>
<td>131</td>
<td>0.313</td>
</tr>
</tbody>
</table>

Model: deviance df intercept mean input prob
609.988 5 -0.201 0.419 0.415

Not significant: Slope, ΔSlope, SlopeComp, R², Dur, PauseN, PauseDur, ArtRate, SpkRate

• Regression analysis for old males alone yields no significant factors or factor groups.

Table 4. Regression analysis for (ð) for young females

<table>
<thead>
<tr>
<th>WORD</th>
<th>p = 0.0051</th>
<th>logodds uncentered weight</th>
<th>tokens</th>
<th>stop/stop+fric</th>
</tr>
</thead>
<tbody>
<tr>
<td>this</td>
<td>0.550</td>
<td>0.65</td>
<td>13</td>
<td>0.462</td>
</tr>
<tr>
<td>the</td>
<td>0.219</td>
<td>0.57</td>
<td>155</td>
<td>0.394</td>
</tr>
<tr>
<td>that</td>
<td>-0.769</td>
<td>0.33</td>
<td>76</td>
<td>0.211</td>
</tr>
<tr>
<td>range</td>
<td>32</td>
<td>(these &amp; those excluded as knockouts)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

ΔSLOPE | p = 0.0261 | (continuous predictor)
| logodds  | 0.822 |
| +1 |

Model: deviance df intercept mean
299.017 4 -0.581 0.34

Not selected as significant: SLOPE, SLOPEComp, R², DUR, PAUSEN, ARTRATE, SPKRATE

• For young females ΔSLOPE is significant, indicating
  ○ (ð) has an increasing likelihood of being stopped as a given segment’s slope becomes steeper than the previous segment...

• The fact that the younger speakers show an effect by lexical item, but the older speakers do not, appears to support the claim (made by Van Herk et al.) that the younger speakers are employing this feature for identity work.

8. RESULTS: Velar nasal fronting (ing)

Table 5. % fronted and Ns for (ing)

<table>
<thead>
<tr>
<th></th>
<th>-ing*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Old Males</td>
<td>84/88</td>
</tr>
<tr>
<td>Young Females</td>
<td>86/176</td>
</tr>
<tr>
<td>totals</td>
<td>170/264</td>
</tr>
</tbody>
</table>

* For fixed effects, Rbrul reports both logodds values and GoldVarb-like factor weights. I report both here. Results for continuous predictors (like SLOPE) cannot be reported or understood in terms of factor weights…
• The high rate of -in’ for the old males makes (ing) less ideal than I had hoped as a comparison test case.
  o Altogether, old males have only 4 full -ing realizations, all nouns and pronouns: everything x2, nothing, and pudding.

Table 6. Mixed-model analysis for (ing) for all speakers

<table>
<thead>
<tr>
<th>AGE+GENDER</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>logodds</td>
</tr>
<tr>
<td>old male</td>
<td>1.671</td>
</tr>
<tr>
<td>young female</td>
<td>-1.671</td>
</tr>
<tr>
<td>range</td>
<td>65</td>
</tr>
</tbody>
</table>

Table 7. Regression analysis for (ing) for young females

<table>
<thead>
<tr>
<th>GRAMMATICAL CAT</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>logodds</td>
</tr>
<tr>
<td>progressive</td>
<td>1.341</td>
</tr>
<tr>
<td>gerund-part</td>
<td>0.127</td>
</tr>
<tr>
<td>gerund</td>
<td>-0.199</td>
</tr>
<tr>
<td>noun</td>
<td>-0.328</td>
</tr>
<tr>
<td>adjective</td>
<td>-0.942</td>
</tr>
</tbody>
</table>

range: 51

SLOPE

<table>
<thead>
<tr>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>logodds</td>
</tr>
<tr>
<td>+1</td>
</tr>
</tbody>
</table>

Model:

deviance | df | intercept | mean
208.887 | 6   | 0.155 | 0.489

Not selected as significant:

PRECEDING ENV, FOLLOWING ENV, ΔSLOPE, SLOPECOMP

Not submitted to run:

R², DUR, PAUSEN, PAUSEDUR, ARTRATE, SPKRATE

• In comparison to, say Hazen 2008, the linguistic results are fairly expected:
  o Progressives and Gerund-Participles (e.g. ...without getting caught) favor -in’, others disfavor -in’
    - Table 6: Progressive (.72) > Gerund-Participle (.54) > Gerund (.38) > Noun (.28) > Adjective (.21)
  o Note the low token count for gerund-participles in Table 7, this may account for the slight disfavoring effect in the young females’ data.

• But what about the Henderson Graph metrics?
  o The SLOPE has significant and sizeable disfavoring effect:
    - As SLOPE increases the likelihood of an -ing realization increases.

- Near categorical use of -in’ by the old males makes logistic regression impossible &/or pointless.
  o (However, as we might expect, grammatical category is significant for runs for the old males – only nouns allow the possibility of -ing.)
9. DISCUSSION

• So, what’s this mean for my questions & hypotheses?
  o For (δ), we find that ΔSLOPE is significant.
    ▪ Young females appear to be actively making use of (δ).
    ▪ This appears to support Labov’s (1972) claim that paralinguistic cues can lend insight into stylistic variation.
  o (ing) was supposed to be a control variable. It was not expected to be locally salient.
    ▪ Due to the near categorical use of -in’ by the old males, can’t really compare the groups here.
    ▪ It’s likely that (ing) patterns with SLOPE in similar fashion for both groups.

• The effect of SLOPE on (ing) appears promising in general.
  o E.g., Guy (1980), studying t/d deletion, raised an interest in the “rate of speech” and said that the “probability of deletion apparently increases in proportion to the rate of speech” (1980: 9) but noted that there wasn’t yet a reliable way to measure speech rate, so didn’t include it in his analysis.
  o I haven’t looked at t/d deletion, but have presented here a method that could test Guy’s claim about speech rate, and related variables.
  o Note that ART_RATE (“rate of speech” in σ/second) is not significant anywhere here, but SLOPE and ΔSLOPE are.
  o The Henderson Slope, in fact, by compounding a number of sequential temporal components of speech may actually be providing us a more perceptually real measure of “rate of speech”…

• These results are promising, but surely more testing is needed…

10. CONCLUSION

• The Henderson Graph’s slope metric appears promising:
  o As a way to study paralinguistic cues to attention-to-speech (Labov 1972).
  o As a way to explore the relationship between language variation, performance, and speech processing.

11. SELECTED REFERENCES


I gratefully thank: Gerard Van Herk and Becky Childs for sharing their data with me, and Gerard Van Herk doubly for both inspiring and humoring this project; Daniel Ezra Johnson for making Rbrul available and for generously answering my many questions about it and about statistical methods and theory more generally; Walt Wolfram and the William C. Friday Endowment at NCSU for continuing support for my work on The Sociolinguistic Archive and Analysis Project.