Toward a variationist, sociolinguistic analysis of pause

1. What is a sociolinguistic analysis of pause?

“One of the most common functions of discourse is to communicate something, but the proper study of linguistics is not communication. (In this case I agree with Chomsky.) Linguists are concerned with the use of language in communication, but that is a very different thing. To take an obvious example, conversation analysts […] and psychologists […] have shown the significance of pauses and silence in communicating. However, there can be no linguistic analysis of silence, though pauses may be a guide to linguistic units.” (Macaulay 2002: 284)

2. Pause as a “linguistic variable”

- “The notion of the linguistic variable has been applied to a full range of levels within language, though not without some argument about its appropriate application. Phonological and morphosyntactic variation have tended to dominate language variation research in synchronic and apparent time studies while syntactic variation has been a significant locus of investigation in diachronic studies” (Wolfram 2006: 335).
- “The empirical reality is that the boundaries of significant and insignificant language variation are often gradient and obscure rather than discrete and transparent” (Wolfram 2006: 334).

Methods and Data

3. The North Carolina Sociolinguistic Archive and Analysis Project

Read more about NC SLAAP at: http://ncslaap.lib.ncsu.edu/

- Recordings are digitized and then transcribed in Praat using TextGrids
  - Transcript lines are time-stamped to a high-degree of accuracy
  - A line of text corresponds to a phonetic utterance
  - A pause is a blank line (with start and end times)

- Audio files and transcripts are uploaded into the NC SLAAP web-based software

*Research reported here is supported by the North Carolina State University Libraries and the William C. Friday Endowment at NC State University. Morphosyntactic and phonological data are from Rowe (2005), Rowe and Kendall (2004), D’Andrea (2005), and Kendall and Wolfram (2006) (research supported by the National Science Foundation Grant BCS-0236838). Thanks to my colleagues for generously sharing these data. Special thanks also to Kirk Hazen and Christine Mallinson for the use of their interviews.
4. Count vs. don’t count forms

- Turn-internal pauses only
e.g.,

453 FW:  ... she thought most of them stayed around here but
454  [ don’t count: 221 ms ]
455 G:  I think most of them did
456 G:  [ count: 222 ms ]
457 G:  but now my children left
458  [ don’t count:: 42 ms ]
459 FW:  yeah

- Thresholds
  o Only counting pauses longer than 60 ms (low threshold)
  o No high threshold

5. Measuring pause

- Arithmetic mean is unreliable for pause
  o Using median as a better indicator of central tendency.
- NC SLAAP analysis feature conducts the count in real-time (Fig. 3)
  o Generates
    ▪ (Basic) statistics
    ▪ Charts (a work in progress)
    ▪ Overviews of the results

Figure 3: NC SLAAP Pause analysis screen for Town Manager

Putative Analysis

6. African Americans, median pause duration from sociolinguistic interviews

<table>
<thead>
<tr>
<th>Name</th>
<th>Gender</th>
<th>YOB</th>
<th>Region</th>
<th>Average Pause Duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Karen</td>
<td>Female</td>
<td>1916</td>
<td>Western NC</td>
<td>334 ms</td>
</tr>
<tr>
<td>Gail Ann</td>
<td>Female</td>
<td>1932</td>
<td>Western NC</td>
<td>363 ms</td>
</tr>
<tr>
<td>Yvonne</td>
<td>Female</td>
<td>1971</td>
<td>Western NC</td>
<td>557 ms</td>
</tr>
<tr>
<td>Mayor</td>
<td>Female</td>
<td>1964</td>
<td>Central NC</td>
<td>333 ms</td>
</tr>
<tr>
<td>Manager</td>
<td>Male</td>
<td>1948</td>
<td>Central NC</td>
<td>322 ms</td>
</tr>
<tr>
<td>Commissioner</td>
<td>Female</td>
<td>1941</td>
<td>Eastern NC</td>
<td>371 ms</td>
</tr>
</tbody>
</table>

7. European Americans, median pause duration from sociolinguistic interviews

All from Warren County (North-Central NC)

<table>
<thead>
<tr>
<th>Name</th>
<th>Gender</th>
<th>YOB</th>
<th>Average Pause Duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>EBH</td>
<td>Male</td>
<td>1952</td>
<td>303 ms</td>
</tr>
<tr>
<td>SM</td>
<td>Male</td>
<td>1975</td>
<td>472 ms</td>
</tr>
<tr>
<td>LW</td>
<td>Female</td>
<td>1983</td>
<td>372 ms</td>
</tr>
</tbody>
</table>
8. Excerpt from SM’s transcript

[As retrieved from NC SLAAP (with very slight editing). Superscript numbers are transcript line numbers]

SM: 82 mm-hmm [pause 0.07] 84 it took six hours

FW: 86 six hours to do three songs

SM: 88 well not cause of playing but it was [pause 0.07] 90 uh [pause 0.30] 92 we’d lay in [pause 0.01] 94 all the rhythm [pause 0.55] 96 all the music [pause 0.55] cause there weren’t really any leads on it

FW: 101 mm-hmm

SM: 104 so [pause 0.11] cause you have to do each thing in different times [pause 0.59] we don’t have to but that’s the way this place does it [pause 0.54] I mean [pause 0.09] they know what they’re doing there [pause 0.66] and then the vocals tracks come in [pause 0.67] so it took like [pause 2.27] plus all the settin’ up and everything I mean you’re on the clock then [pause 0.03] so

FW: 123 [mm-hmm]

9. Comparison of pause with diagnostic variables in two speech situations for Town Manager

<table>
<thead>
<tr>
<th></th>
<th>Plural →s abs</th>
<th>Cop abs</th>
<th>3rd sg. →s Abs</th>
<th>Past tense be Reg.</th>
<th>Pre-V CCR</th>
<th>Post-V r-lessness</th>
<th>Median Pause Dur</th>
</tr>
</thead>
<tbody>
<tr>
<td>Radio Interview</td>
<td>4/24</td>
<td>5/9</td>
<td>5/16</td>
<td>5/7</td>
<td>5/11</td>
<td>42/68</td>
<td>300 ms</td>
</tr>
<tr>
<td></td>
<td>5/28</td>
<td>(4/7 are)</td>
<td>55.7%</td>
<td>62.5%</td>
<td>57.1%</td>
<td>45.5%</td>
<td></td>
</tr>
<tr>
<td>Sociolinguistic Interview</td>
<td>19/96</td>
<td>7/31</td>
<td>6/10</td>
<td>26/36</td>
<td>30/57</td>
<td>46/100</td>
<td>322 ms</td>
</tr>
<tr>
<td></td>
<td>19.8%</td>
<td>(3/7 are)</td>
<td>22.6%</td>
<td>60%</td>
<td>72.2%</td>
<td>52.6%</td>
<td></td>
</tr>
</tbody>
</table>

10. Comparison of pause with diagnostic variables in two speech situations for Mayor

<table>
<thead>
<tr>
<th></th>
<th>Plural →s abs</th>
<th>Cop abs</th>
<th>3rd sg. →s Abs</th>
<th>Past tense be Reg.</th>
<th>Pre-V CCR</th>
<th>Post-V r-lessness</th>
<th>Median Pause Dur</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speech at Town Event</td>
<td>14/28</td>
<td>6/13</td>
<td>4/6 are</td>
<td>10/18</td>
<td>0/0</td>
<td>7/9</td>
<td>39/93 500 ms</td>
</tr>
<tr>
<td></td>
<td>50.0%</td>
<td>46.2%</td>
<td>55.6%</td>
<td>NA</td>
<td>NA</td>
<td>77.8%</td>
<td></td>
</tr>
<tr>
<td>Sociolinguistic Interview</td>
<td>30/94</td>
<td>7/12</td>
<td>4/4 are</td>
<td>17/18</td>
<td>15/24</td>
<td>45/66</td>
<td>51/100 333 ms</td>
</tr>
<tr>
<td></td>
<td>31.9%</td>
<td>58.3%</td>
<td>94.4%</td>
<td>62.5%</td>
<td>68.2%</td>
<td>51.0%</td>
<td></td>
</tr>
</tbody>
</table>
11. What Next?

- Impressions are that a larger study will find significance with social variables
  - Style
  - Community of Practice?
  - Other “primary” independent variables (e.g., gender)?

- Need (much) more data
  - I’m continuing to transcribe and digitize
  - An advantage with NC SLAAP is that all data that is added to the archive becomes available (with the PI’s permission) immediately for this (and other) analyses

- I’m continuing to develop the analysis tool
  - POS Tagging for pause location analysis
  - Refining count algorithm, statistics, graphic generation,…

- More than this particular analysis/question, I’m interested in developing tools (NC SLAAP features) to aid in new analyses and the testing of new hypotheses

12. Selected References


Mukherjee, Joybrato. (2000) Speech is Silver, but Silence is Golden: Some Remarks on the Function(s) of Pauses”, Anglia 118: 571-84.


