Computational analysis of AphasiaBank transcripts and video

**Transcription**

Interviews are transcribed in CHAT (MacWhinney, 2000) and linked to the video. The methods for doing this emphasize the production of high-quality transcriptions that can be checked for accuracy and reliability. Both are explained below.

**MOR**

All possible parts of speech appear for each lexical item.

- **CLAN command (verb)**
  
  1. Post: all parts of speech are distinguished based on context before and after the word.
  
  2. **POST**

**Errors**

An error coding system (available at the AphasiaBank website) was described to capture errors at the word and utterance levels. After coding errors in the transcription, you can do a wide range of error analyses. There are 2 examples of CLAN commands analyzing all word level errors for the 3 AphasiaBank transcripts.

- **CLAN command (verb)**
  
  1. Post: all parts of speech are distinguished based on context before and after the word.
  
  2. **POST**

**Timelocked Display**

Linking between the transcription and the video or audio record is done by replaying the recording (in PS format) and pressing the speaker at the end of each utterance. This automatically inserts "halts" at the end of the utterance, which correspond to where another speaker appears in the recording. Thus, during playback, CLAN highlights the currently playing utterance.

**Linked-In Video**

The CHAT format can easily be converted to ELAN using the annotations on video and audio resources. Transcripts in ELAN is a professional tool for the creation of complex annotations.

CLAN command:

- **Temporal display (verb)**

**MORTALITY**

- **Creates an Excel spreadsheet listing frequencies for all parts of speech and bound morphemes**

**CLAN command (verb)**

<table>
<thead>
<tr>
<th>Gender</th>
<th>Aphasia type</th>
<th>File name</th>
<th>Tokens</th>
<th>Errors</th>
<th>Poses</th>
<th>% errors</th>
<th>% poses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>Broca</td>
<td>file002a</td>
<td>10</td>
<td>11</td>
<td>1</td>
<td>9.0</td>
<td>90.0</td>
</tr>
<tr>
<td>Female</td>
<td>Wernicke</td>
<td>file001a</td>
<td>11</td>
<td>13</td>
<td>1</td>
<td>8.0</td>
<td>92.0</td>
</tr>
</tbody>
</table>

**Errors**

General communication can be studied through the method of testing files, in which shaded coding files maintain their linkages to segments of the video.

**Gestures**

File: transcript.txt

<table>
<thead>
<tr>
<th>Action</th>
<th>Segment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Left hand crosses back to left arm</td>
<td>101, 102</td>
</tr>
<tr>
<td>Front</td>
<td>Gave toy to alderella, from classification action description Meaning: Repeating and teaching dress names</td>
</tr>
</tbody>
</table>

For detailed guidelines, visit: http://elab.org/aphasiabank/