Lecture 12: Some Medical Applications: Intoxication, Depression, Trauma
Topic 1: Intoxication
Hollien et al 2001

- **Methods:**
  - 35 young adults, 19 males, 16 females
  - given series of doses of alcohol
  - speech collected at 4 BAC stages
    - Rainbow passage
    - difficult words (buttercup, shapupie)
    - extemp speech (“Tell us about your favorite TV program)
    - head-mounted mikes

- **Investigated:**
  - F0 mean and variance
  - duration/rate of speech
  - intensity
  - disfluencies
Hollien et al 2001 Results:
F0

![Graph showing the relationship between SFF shifts in ST and BrAC for males and females.](image)
# Hollien et al 2001 Results: Duration

<table>
<thead>
<tr>
<th>Group</th>
<th>Level of intoxication (BrAC)</th>
<th>Shift (0.00–0.12)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.00</td>
<td>0.04</td>
</tr>
<tr>
<td><strong>Men</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (s)</td>
<td>25.3</td>
<td>25.8</td>
</tr>
<tr>
<td>S.D. (s)</td>
<td>2.9</td>
<td>2.5</td>
</tr>
<tr>
<td><strong>Women</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (s)</td>
<td>25.1</td>
<td>25.5</td>
</tr>
<tr>
<td>S.D. (s)</td>
<td>2.2</td>
<td>2.2</td>
</tr>
</tbody>
</table>
### Hollien et al 2001 Results: Disfluencies

<table>
<thead>
<tr>
<th>Subjects</th>
<th>N</th>
<th>(0.00)</th>
<th>(0.04)</th>
<th>(0.08)</th>
<th>(0.12)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Males</td>
<td>19</td>
<td>3.2</td>
<td>4.7</td>
<td>6.5</td>
<td>8.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.0</td>
<td>2.6</td>
<td>3.1</td>
<td>3.4</td>
</tr>
<tr>
<td>Females</td>
<td>16</td>
<td>2.2</td>
<td>3.5</td>
<td>4.7</td>
<td>6.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.7</td>
<td>2.2</td>
<td>2.7</td>
<td>3.0</td>
</tr>
<tr>
<td>Mean</td>
<td>35</td>
<td>2.7</td>
<td>4.1</td>
<td>5.6</td>
<td>7.4</td>
</tr>
</tbody>
</table>
Hollien et al
2001 Results:
Magnitudes
Hollien et al 2001 Results: Speaker Specific Effects

- 20% of speakers did not follow these trends
A famous case study

The *Exxon Valdez* oil spill occurred in Prince William Sound, Alaska, on March 24, 1989, when the *Exxon Valdez*, an oil tanker bound for Long Beach, California, struck Prince William Sound's Bligh Reef and spilled 260,000 to 750,000 barrels (41,000 to 119,000 m³) of crude oil.[1][2] It is considered to be one of the most devastating human-caused environmental disasters.[3] As
Was Captain Hazelwood drunk?

- Not clear if this is relevant, since seems like other questionable corporate things were going on:
  - he was asleep below deck
  - The third mate was in charge of the wheelhouse
  - the ship’s radar was broken
- But is a well-studied case
Johnson et al examined 3 kinds of cues

- Segmental Effects
- Disfluencies
- Suprasegmental Effects
Keith Johnson's /s/ and /ʃ/

Fig. 1. Power spectra of /s/ (a) and /ʃ/ (b) produced by K. J. in a quiet recording booth with recording equipment responsive up to 5,000 Hz.
/ʃ/: Captain Hazelwood

Fig. 2. Power spectra of /ʃ/ produced by Captain Hazelwood in the words she's and shout recorded 33 h before the accident. Each spectrum is paired with a spectrum of the background noise from a nearby open-mike pause.
33 Hrs before

1 Hr before

Immediately after

1 Hr after

9 Hrs after
Fig. 2. Power spectra of /ʃ/ produced by Captain Hazelwood in the words she's and shout recorded 33 h before the accident. Each spectrum is paired with a spectrum of the background noise from a nearby open-mike pause.
Duration

Segment Durations of "Exxon Valdez"

- 33 Hrs before
- 1 Hr before
- Immediately after
- 1 Hr after
- 9 Hrs after

Duration (ms)
F0 Measurements

F0 (Hz)

Time of Recording

-33 -1 0 +1 +9 CC

Exxon Valdez
13-16
Interview
Table 3. Summary of phenomena found in the analysis of the NTSB tape (numbers in parentheses indicate the time of recording)

<table>
<thead>
<tr>
<th>Gross effects</th>
<th>revisions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(-1) Exxon Ba, uh Exxon Valdez</td>
</tr>
<tr>
<td></td>
<td>(-1) departed disembarked</td>
</tr>
<tr>
<td></td>
<td>(-1) I, we’ll</td>
</tr>
<tr>
<td></td>
<td>(-1) columbia gla, columbia bay</td>
</tr>
</tbody>
</table>

| Segmental effects             | misarticulation of /r/ and /l/ |
|-------------------------------| (0) northerly, little, drizzle, visibility |
|                               | (/s/ becomes /ʃ/ (fig. 3) |
|                               | final devoicing (e.g. /z/ → /s/) |
|                               | (-1,0,+1) Valdez → Valdes |

| Suprasegmental effects        | reduced speaking rate (fig. 4, 5) |
|-------------------------------| mean change in pitch range |
|                               | (talker-dependent, fig. 6) |
|                               | increased F0 jitter (fig. 6) |
Problems

• If intoxicated speech, why wasn’t s pronounced as sh 1 hour before?
• Other kinds of speaker state could cause drop in F0, slower speech, and disfluencies?
  • Stress, just having woken up, trauma....
Automatic Classification

- Use of prosodic speech characteristics for automated detection of alcohol intoxication
  Michael Levit, Richard Huber, Anton Batliner, Elmar Noeth

- Break utterance into phrases automatically, based on
  - fundamental frequency (where possible);
  - zero-crossing rate

![Waveform Diagram with Phrases PhU1, PhU2, PhU3]
Then use 4 classes of features

- Prosodic
  - F0 max, F0 min, energy max, energy min, pause length
- Duration of voiced regions, unvoiced regions, etc.
- Jitter and shimmer
  - jitter is variation in pitch
  - shimmer is variation in energy
- Average cepstrum and cepstral slope
Methods

- Alcoholized speech samples collected at the Police Academy of Hessen, Germany
- 120 readings (87 minutes) of a fable
- 33 male speakers
- BAC between 0 and .24/mille

<table>
<thead>
<tr>
<th>Alcohol Blood Level</th>
<th>0.0</th>
<th>&lt; 0.4</th>
<th>&lt; 0.8</th>
<th>&lt; 1.2</th>
<th>&lt; 1.6</th>
<th>&lt; 2.0</th>
<th>&lt; 2.4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recordings</td>
<td>32</td>
<td>20</td>
<td>20</td>
<td>18</td>
<td>20</td>
<td>7</td>
<td>3</td>
</tr>
</tbody>
</table>

- Binary task: above or below 0.8/mille
- leave-one-out cross-validation
- neural net classifier
Results of Levit et al.

- Used dev set to find best classifier
- This suggested two feature classes:
  - Prosodic features
  - Jitter/shimmer
- Results with this classifier
  - 62% phrase-accuracy
  - 69% for the whole speech sample
    - voting of the phrases
New Corpus!

- Alcohol Language Corpus
  - Florian Schiel et al 2009, 2010
  - [http://www.bas.uni-muenchen.de/forschung/Bas/BasALCeng.html](http://www.bas.uni-muenchen.de/forschung/Bas/BasALCeng.html)
- 162 speakers (77 female, 85 male)
  - recorded in a car (sometimes with engine running)
  - command and control speech ("turn off the radio")
  - spontaneous dialogue, monologue, question answering
  - read speech
  - counts of disfluencies, etc
- sample, drunk:
- sample, sober:
Automatic detection in ALC: Paralinguistic Challenge 2011

- **Human**: 66-72% (Schiell 2011, Ultes, Schmitt, Minker 2011)
- **Machine**: roughly 65%-70%
- **Example features from winning system:**


- **Prosody** (f0, duration, energy, jitter, shimmer)
- **Spectral** (MFCC, MFB log-energy, formants)
- **Computed over whole utterance and small windows**
- **normalized phoneme duration**
- **iterative speaker normalization**
Topic 2: Depression
Stirman and Pennebaker

- Suicidal poets
- 300 poems from early, middle, late periods of
  - 9 suicidal poets
  - 9 non-suicidal poets
Stirman and Pennebaker: 2 models

- Durkheim disengagement model:
  - suicidal individual has failed to integrate into society sufficiently, is detached from social life
  - detach from the source of their pain, withdraw from social relationships, become more self-oriented
  - prediction:
    - more self-reference, less group references

- Hopelessness model:
  - Suicide takes place during extended periods of sadness and desperation, pervasive feelings of helplessness, thoughts of death
  - prediction:
    - more negative emotion, fewer positive, more refs to death
Methods

- 156 poems from 9 poets who
  - committed suicide
  - published, well-known
  - in English
  - have written within 1 year of committing suicide

- Control poets matched for nationality, education, sex, era.
### TABLE 2. Suicidal Poets and Their Controls

<table>
<thead>
<tr>
<th>Suicidal Poet</th>
<th>Age at Death</th>
<th>Control Poet</th>
<th>Cutoff Age</th>
<th>Nationality</th>
<th>Other Similarities</th>
</tr>
</thead>
<tbody>
<tr>
<td>John Berryman (1914–1972)</td>
<td>58</td>
<td>Lawrence Ferlinghetti (1919–)</td>
<td>59 (1978)</td>
<td>American</td>
<td>PhD</td>
</tr>
<tr>
<td>Adam L. Gordon (1833–1870)</td>
<td>37</td>
<td>Matthew Arnold (1822–1888)</td>
<td>45 (1867)</td>
<td>British</td>
<td></td>
</tr>
<tr>
<td>Sarah Teasdale (1884–1933)</td>
<td>49</td>
<td>Edna St. V. Millay (1892–1950)</td>
<td>49 (1941)</td>
<td>American</td>
<td></td>
</tr>
<tr>
<td>Hart Crane (1899–1932)</td>
<td>33</td>
<td>Joyce Kilmer (1886–1918)</td>
<td>32 (1918)</td>
<td>American</td>
<td></td>
</tr>
<tr>
<td>Sergei Esenin (1895–1925)</td>
<td>30</td>
<td>Boris Pasternak (1890–1960)</td>
<td>35 (1930)</td>
<td>Russian</td>
<td></td>
</tr>
<tr>
<td>Vladimir Mayakovskiy (1893–1930)</td>
<td>37</td>
<td>Osip Mandelstam (1891–1938)</td>
<td>37 (1928)</td>
<td>Russian</td>
<td></td>
</tr>
</tbody>
</table>
TABLE 1. Means for LIWC Categories

<table>
<thead>
<tr>
<th></th>
<th>Suicide Group</th>
<th>Control Group</th>
<th>Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Early</td>
<td>Middle</td>
<td>Later</td>
</tr>
<tr>
<td>Disengagement theory</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I (me, my)</td>
<td>4.0</td>
<td>3.4</td>
<td>4.0</td>
</tr>
<tr>
<td>We (us, our)</td>
<td>.73</td>
<td>1.3</td>
<td>.85</td>
</tr>
<tr>
<td>Communication (talk, share)</td>
<td>1.2</td>
<td>1.1</td>
<td>1.0</td>
</tr>
<tr>
<td>Hopelessness theory</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative emotion (hate, worthless)</td>
<td>2.2</td>
<td>1.8</td>
<td>1.7</td>
</tr>
<tr>
<td>Positive emotion (happy, love)</td>
<td>3.3</td>
<td>3.1</td>
<td>3.9</td>
</tr>
<tr>
<td>Death (dead, grave)</td>
<td>.52</td>
<td>.47</td>
<td>.69</td>
</tr>
<tr>
<td>Other findings</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sexual words (lust, breast)</td>
<td>.60</td>
<td>.84</td>
<td>.47</td>
</tr>
</tbody>
</table>

Note: Means reflect percentage of total words used in each poem within the relevant category. Effects refer to: S = suicide vs. nonsuicide main effect, P = phase of career main effect. All effects are significant p ≤ .05, except ** p ≤ .08.
Significant factors

- Disengagement theory
  - I, me, mine
  - we, our, ours
- Hopelessness theory
  - death, grave
- Other
  - sexual words (lust, breast)
Beck (1967) cognitive theory of depression
- depression-prone individuals see the world and themselves in pervasively negative terms

Pyszynski and Greenberg (1987)
- think about themselves
- after the loss of a central source of self-worth, unable to exit a self-regulatory cycle concerned with efforts to regain what was lost.
- results in self-focus, self-blame

Durkheim social integration/disengagement
- perception of self as not integrated into society is key to suicidality and possibly depression
Methods

- College freshmen
  - 31 currently-depressed (standard inventories)
  - 26 formerly-depressed
  - 67 never-depressed
- Session 1: take depression inventory
- Session 2: write essay
  - please describe your deepest thoughts and feelings about being in college... write continuously off the top of your head. Don’t worry about grammar or spelling. Just write continuously.
Results

- depressed used more “I, me” than never-depressed
  - turned out to be only “I”
- and used more negative emotional words
- not enough “we” to check Durkheim model
- formerly depressed participants used more “I” in the last third of the essay
Study 1: Use LIWC counts on posts from 320 English and Spanish forums
  - 80 posts each from depression forums in English and Spanish
  - 80 control posts each from breast cancer forums

Run the following LIWC categories
  - I
  - we
  - negative emotion
  - positive emotion
## Results of Study 1

<table>
<thead>
<tr>
<th>Categories</th>
<th>English</th>
<th></th>
<th>Spanish</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Dep.</td>
<td>Breast Cancer</td>
<td>Dep.</td>
<td>Breast Cancer</td>
</tr>
<tr>
<td></td>
<td>N=80</td>
<td></td>
<td>N=80</td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>Mean</td>
<td>Mean</td>
<td>Mean</td>
<td>Mean</td>
</tr>
<tr>
<td>(SD)</td>
<td>(SD)</td>
<td>(SD)</td>
<td>(SD)</td>
<td>(SD)</td>
</tr>
<tr>
<td>First person singular</td>
<td>12.24</td>
<td>4.03</td>
<td>9.30</td>
<td>5.03</td>
</tr>
<tr>
<td></td>
<td>(.297)</td>
<td>(.301)</td>
<td>(.234)</td>
<td>(.276)</td>
</tr>
<tr>
<td>First person plural</td>
<td>.18</td>
<td>.72</td>
<td>.22</td>
<td>1.02</td>
</tr>
<tr>
<td></td>
<td>(.33 )</td>
<td>(1.06)</td>
<td>(.39 )</td>
<td>(1.28)</td>
</tr>
<tr>
<td>Positive Emotions</td>
<td>1.72</td>
<td>2.54</td>
<td>2.99</td>
<td>3.53</td>
</tr>
<tr>
<td></td>
<td>(1.14)</td>
<td>(1.72)</td>
<td>(1.36)</td>
<td>(1.93)</td>
</tr>
</tbody>
</table>
Study 2

- From depression forums:
  - 404 English posts
  - 404 Spanish posts
- Create a term by document matrix of content words
  - 200 most frequent content words
- Do a factor analysis
  - dimensionality reduction in term-document matrix
  - Used 5 factors
# English Factors

<table>
<thead>
<tr>
<th>FACTOR 1: Treatment</th>
<th>FACTOR 1: Disclosure</th>
<th>FACTOR 3: Family</th>
<th>FACTOR 4: Symptoms</th>
<th>FACTOR 5: School</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medication</td>
<td>.62</td>
<td>.43</td>
<td>Mom</td>
<td>.49</td>
</tr>
<tr>
<td>Effect</td>
<td>.46</td>
<td>People</td>
<td>.41 Daughter</td>
<td>.48</td>
</tr>
<tr>
<td>Depression</td>
<td>.43</td>
<td>Know</td>
<td>.39 Child</td>
<td>.48</td>
</tr>
<tr>
<td>Side</td>
<td>.35 Happy</td>
<td>.35</td>
<td>Family</td>
<td>.48</td>
</tr>
<tr>
<td>Week</td>
<td>.34 Talk</td>
<td>.35 Brother</td>
<td>.43</td>
<td></td>
</tr>
<tr>
<td>Therapy</td>
<td>.34 Feel</td>
<td>.34 Sister</td>
<td>.43</td>
<td></td>
</tr>
<tr>
<td>Suffer</td>
<td>.34 Want</td>
<td>.34 Dad</td>
<td>.41</td>
<td></td>
</tr>
<tr>
<td>Disorder</td>
<td>.33 Suppose</td>
<td>.33 Son</td>
<td>.33</td>
<td></td>
</tr>
<tr>
<td>Doctor</td>
<td>.33 Read</td>
<td>.32 Love</td>
<td>.33</td>
<td></td>
</tr>
<tr>
<td>Antidepressant</td>
<td>.32 Hurt</td>
<td>.32 Girl</td>
<td>.33</td>
<td></td>
</tr>
<tr>
<td>Experience</td>
<td>.32 Wrong</td>
<td>.32 Young</td>
<td>.32</td>
<td></td>
</tr>
<tr>
<td>Major</td>
<td>.32 Emotional</td>
<td>.31 Parent</td>
<td>.32</td>
<td></td>
</tr>
<tr>
<td>Mental</td>
<td>.31 Mind</td>
<td>.31 House</td>
<td>.31</td>
<td></td>
</tr>
<tr>
<td>Psychiatrist</td>
<td>.31 Sad</td>
<td>.31 Husband</td>
<td>.30</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Make</td>
<td>.31 Crazy</td>
<td>.30</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>FACTOR 4: Symptoms</th>
<th>FACTOR 5: School</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sleep</td>
<td>.51 Constant</td>
</tr>
<tr>
<td>Hour</td>
<td>.48 Relationship</td>
</tr>
<tr>
<td>Food</td>
<td>.44 School</td>
</tr>
<tr>
<td>Wake</td>
<td>.44 High</td>
</tr>
<tr>
<td>Morning</td>
<td>.44 Lack</td>
</tr>
<tr>
<td>Night</td>
<td>.41 University</td>
</tr>
<tr>
<td>Bed</td>
<td>.39 Social</td>
</tr>
<tr>
<td>Stay</td>
<td>.38 College</td>
</tr>
<tr>
<td>Weight</td>
<td>.37 Move</td>
</tr>
<tr>
<td>Eat</td>
<td>.36 Friend</td>
</tr>
<tr>
<td>Place</td>
<td>.32 Girlfriend</td>
</tr>
<tr>
<td></td>
<td>Class</td>
</tr>
</tbody>
</table>
## Spanish Factors

<table>
<thead>
<tr>
<th>FACTOR 1: Family</th>
<th>FACTOR 2: Relationship History</th>
</tr>
</thead>
<tbody>
<tr>
<td>MADRE/mother</td>
<td>RELACION/relationship</td>
</tr>
<tr>
<td>HERMANO/brother</td>
<td>ENAMORADO/love</td>
</tr>
<tr>
<td>ABUELA/grandmother</td>
<td>CONOCI/met</td>
</tr>
<tr>
<td>PADRES/parents</td>
<td>HABLAR/talk</td>
</tr>
<tr>
<td>PAPA/father</td>
<td>CHICO/guy</td>
</tr>
<tr>
<td>HORRIBLE/horrible</td>
<td>AMIGOS/friends</td>
</tr>
<tr>
<td>CASA/house</td>
<td>NOVIO/boyfriend</td>
</tr>
<tr>
<td>SUICIDIO/suicide</td>
<td>JUNTOS/together</td>
</tr>
<tr>
<td>DINERO/money</td>
<td>ESPECIAL/special</td>
</tr>
<tr>
<td>ESTUDIOS/studies</td>
<td>EMPECE/start</td>
</tr>
<tr>
<td>CLASE/class</td>
<td>TIEMPO/time</td>
</tr>
<tr>
<td>FEA/ugly</td>
<td>DEJAR/leave</td>
</tr>
<tr>
<td>ASCO/disgust</td>
<td>FINAL/end</td>
</tr>
<tr>
<td>COMER/eat</td>
<td>HISTORIA/history</td>
</tr>
<tr>
<td>PEQUENIOO/small</td>
<td>MESES/months</td>
</tr>
<tr>
<td>FAMILIA/family</td>
<td>LLEGAR/arrive</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>FACTOR 3: Hopelessness</th>
<th>FACTOR 4: School</th>
</tr>
</thead>
<tbody>
<tr>
<td>NOCHE/night</td>
<td>TIMIDA/timid</td>
</tr>
<tr>
<td>SEGUNDO/second</td>
<td>CARRERA/career</td>
</tr>
<tr>
<td>MORIR/die</td>
<td>COLEGIO/school</td>
</tr>
<tr>
<td>PAZ/peace</td>
<td>CONFIANZA/trust</td>
</tr>
<tr>
<td>OJOS/eyes</td>
<td>ESTUDIOS/studies</td>
</tr>
<tr>
<td>ESPERO/hope</td>
<td>INCAPAZ/incapable</td>
</tr>
<tr>
<td>TERRIBLE/terrible</td>
<td>UNIVERSI/university</td>
</tr>
<tr>
<td>FUERTE/strong</td>
<td>TONTERIA/foolishness</td>
</tr>
<tr>
<td>CORAZON/heart</td>
<td></td>
</tr>
<tr>
<td>SUENIOS/dreams</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>FACTOR 5: Treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSICOLOGO/psychologist</td>
</tr>
<tr>
<td>ANSIEDAD/anxiety</td>
</tr>
<tr>
<td>EMPRESA/company</td>
</tr>
<tr>
<td>ANTIDEP/antidepressants</td>
</tr>
<tr>
<td>SINTOMAS/symptoms</td>
</tr>
<tr>
<td>MEDICAMENTO/medicines</td>
</tr>
</tbody>
</table>
Speech features for Depression


Commonly used features:
- F0 variance (monopitch)
- loudness variance (monoloudness)
- rate of speech (slower)
  - response delay, pauses
- spectral features
Social stage model of collective coping (Pennebaker & Harber, 1993). After a traumatic experience:

Stage 1: people cope by sharing their thoughts about the upsetting experience
Stage 2, a few weeks later: decrease in talking, but still thinking about event
Stage 3: 6-8 weeks later: reduction in both talking and thinking

What are the linguistic characteristics of stage 1?
Cohn, Mehl, Pennebaker: Linguistic Markers of Psychological Change Surrounding September 11, 2001

- 1084 LiveJournal users
- all blog entries for 2 months before and after 9/11
- Lumped prior two months into one “baseline” corpus.
- Investigated changes after 9/11 compared to that baseline
- Using LIWC categories
Variables examined

- Emotional positivity
  - difference between LIWC scores for positive emotion words (happy, good, nice) and negative emotion words (kill, ugly, guilty).

- cognitive processing
  - think, question, because: concerned with organizing and intellectually understanding issues

- social orientation
  - talk, share, friends and personal pronouns besides I/me. (essentially counts # of references to other people)
Last factor: Psychological Distancing

• psychological distancing
  • factor-analytic:
    • + articles,
    • + words > 6 letters long
    • - I/me/mine
    • - would/should/could
    • - present tense verbs
  • low score = personal, experiential lg, focus on here and now
  • high score: abstract, impersonal, rational tone
Results

A. Emotional Positivity

B. Cognitive Processes

C. Social Processes

D. Psychological Distancing
LiveJournal.com September 11, 2001 study: Positive and negative emotion words

Livejournal.com:

I, me, my on or after Sep 11, 2001


Graph from Pennebaker slides
September 11 LiveJournal.com study:
We, *us, our*


Graph from Pennebaker slides
Trauma after Princess Diana’s death

- Princess Diana died August 30, 1997
- Over the next 4 weeks, scraped all conversations from “The UK Experience” chat room on AOL.
- 121 chat sessions among 3,139 participants.
- Compared to baseline rates:
  - Increase in *we*
  - Decrease in *I*
  - Increase in negative emotional words
Texas A&M Bonfire tragedy

- Gortner and Pennebaker
- Examined student newspaper in the weeks after the tragedy:
  - Increase in we
  - Increase in I
  - Increase in negative emotion
Another domain of trauma?
Restaurant Reviews

Jurafsky, Chahuneau, Routledge, Smith 2014.

- 6562 restaurants
  - 900K reviews [www.yelp.com](http://www.yelp.com)

- Negative (★):
  - The bartender... absolutely horrible... we waited 10 min before we even got her attention... and then we had to wait 45 - FORTY FIVE! - minutes for our entrees... stalk the waitress to get the cheque... she didn't make eye contact or even break her stride to wait for a response...
What makes a bad review bad?

- Negative sentiment language
  - horrible awful terrible worst bad disgusting
- narrative
  - past tense
    - waited, didn’t make eye contact, was disappointing.
  - 3rd person pronouns
    - he she his her
- other people
  - manager, customer, minutes, money, waitress, waiter, bill, attitude, management, business, apology, mistake, table, charge, order, hostess,
- mentions of we and us
  - we waited 10 min before we even got her attention... and then we had to wait 45 - FORTY FIVE! - minutes for our entrees... ...
We just saw texts with these characteristics!

- Negative sentiment, past tense narratives about others
- Enormous increase in “we” and “us”: solace in community
- Chat group discussions after Princess Diana’s death
- Blog posts after September 11, 2001
- Student newspaper reports after a campus tragedy

• Conclusion: **Awful reviews are trauma narratives**
Speech features for trauma?
Lecture 11b: Disfluency
Outline

Disfluencies
Characteristics of disfluencies
Detecting disfluencies
Fragments
Disfluencies: standard terminology (Levelt)

**Reparandum**: thing repaired

**Interruption point** (IP): where speaker breaks off

**Editing phase** (edit terms): uh, I mean, you know

**Repair**: fluent continuation

---

Does American airlines offer any one–way flights [uh] one–way fares for 160 dollars?
Why disfluencies?

- Need to clean them up to get understanding
  
  Does American airlines offer any one-way flights [uh] one-way fares for 160 dollars?
  
  Delta leaving Boston seventeen twenty one one arriving Fort Worth twenty two twenty one forty

- Might help in language modeling
  
  - Disfluencies might occur at particular positions (boundaries of clauses, phrases)

- Annotating them helps readability

- Disfluencies cause errors in nearby words
Counts (from Shriberg, Heeman)

- Sentence disfluency rate
  - 6% (ATIS human-computer)
  - 10% (ATIS long sentences human-comp)
  - 34% of sentences disfluent (Levelt human-human)
  - 50% (swbd multiword sentences)

- Word disfluency rate
  - 0.4% (ATIS)
  - 6% (SWBD)
  - 13% (AMEX Human-human travel)
Disfluency facts

- Fragments are good cues to disfluencies (Nakatani and Hirschberg 1994)
- The repair often has same structure as reparandum (Hindle 1983)
  - E.g., both are Noun Phrases (NPs) in this example:
    - So if could automatically find IP, could find and correct reparandum!
Um and uh: should we just remove them? No.

- **Different meanings** (Clark and Fox Tree)
  - **uh**: used to announce minor delays
    - Preceded and followed by shorter pauses
  - **um**: used to announce major delays
    - Preceded and followed by longer pauses

- **Different connotations**
  - **um** is used far less often by women (Acton 2012)

- **Removing hurts LM performance**! (Stolcke and Shriber 1996)
  - **Um** and **uh** help predict next word because tend to occur at clause boundaries:
    - Some of the different things we’re doing [uh] there’s not time to do it all
      - (Unless we somehow remove only medial filled pauses)
Goldwater study

- Study error rate in two recognizers
  - SRI/ICSI/UW RT-04 CTS system (Stolcke et al., 2006)
  - CU-HTK RT-04 CTS system (Evermann 2004a, 2005)
- On NIST RT-03 development set.
  - 36 telephone conversations, 72 speakers, 38477 reference words.
- Metric: Individual Word Error Rate:

REF: *** YEAH I would HAVE never BEEN ABLE TO GET it
HYP: YOU KNOW I would **** never **** **** REALLY GOT it
Eval: I S D D D S S
# Coding disfluencies

<table>
<thead>
<tr>
<th>Disfluency</th>
<th>Coding</th>
</tr>
</thead>
<tbody>
<tr>
<td>yeah</td>
<td>Before Rep</td>
</tr>
<tr>
<td>i</td>
<td>First Rep</td>
</tr>
<tr>
<td>i</td>
<td>Middle Rep</td>
</tr>
<tr>
<td>i</td>
<td>Last Rep</td>
</tr>
<tr>
<td>think</td>
<td>After Rep</td>
</tr>
<tr>
<td>you</td>
<td></td>
</tr>
<tr>
<td>should</td>
<td>Before FP</td>
</tr>
<tr>
<td>um</td>
<td></td>
</tr>
<tr>
<td>ask</td>
<td>After FP</td>
</tr>
<tr>
<td>for</td>
<td></td>
</tr>
<tr>
<td>the</td>
<td>Before Frag</td>
</tr>
<tr>
<td>ref-</td>
<td></td>
</tr>
<tr>
<td>recommendation</td>
<td>After Frag</td>
</tr>
</tbody>
</table>
Other factors and (independent) results

- Word class
  - Open
  - Function
  - Discourse marker
- Turn-initial
- Male
- Female
- Word length in phones
- LM unigram, trigram
- Prosodic features

<table>
<thead>
<tr>
<th>Feature</th>
<th>IWER</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>19.8</td>
</tr>
<tr>
<td>Female</td>
<td>16.7</td>
</tr>
<tr>
<td>Starts turn</td>
<td>21.0</td>
</tr>
<tr>
<td>Before FP</td>
<td>16.7</td>
</tr>
<tr>
<td>After FP</td>
<td>16.8</td>
</tr>
<tr>
<td>Before frag</td>
<td>32.2</td>
</tr>
<tr>
<td>After frag</td>
<td>22.0</td>
</tr>
<tr>
<td>Before rep</td>
<td>19.6</td>
</tr>
<tr>
<td>After rep</td>
<td>15.3</td>
</tr>
<tr>
<td>Non-final rep</td>
<td>28.4</td>
</tr>
<tr>
<td>Final rep</td>
<td>12.8</td>
</tr>
<tr>
<td>Open class</td>
<td>17.3</td>
</tr>
<tr>
<td>Closed class</td>
<td>19.3</td>
</tr>
<tr>
<td>Discourse marker</td>
<td>18.1</td>
</tr>
<tr>
<td>All words</td>
<td>18.2</td>
</tr>
</tbody>
</table>
Coefficient values in joint regression

- Left of dotted line -> reduces error
- Right of dotted line -> increases error
Recent work: EARS Metadata Evaluation (MDE)

- A multiyear DARPA bakeoff about a decade ago
- Edit word detection:
  - Find all words in reparandum (words that will be removed)
- Filler word detection
  - Filled pauses (uh, um)
  - Discourse markers (you know, like, so)
  - Editing terms (I mean)
- Interruption point detection

Liu et al 2003
State of the art: Edit word detection

- Multi-stage model
  - HMM combining LM and decision tree finds IP
  - Heuristics rules find onset of reparandum
  - Separate repetition detector for repeated words

- One-stage model
  - CRF jointly finds edit region and IP
  - BIO tagging (each word has tag whether is beginning of edit, inside edit, outside edit)

- Error rates:
  - 43-50% using transcripts
  - 80-90% using ASR

- Problem is still quite unsolved!
Fragments

- Incomplete or cut-off words:
  - Leaving at seven fif- eight thirty
  - uh, I, I d-, don't feel comfortable
  - You know the fam-, well, the families
  - I need to know, uh, how- how do you feel...
  - Uh yeah, yeah, well, it- it- that’s right. And it-
Fragment glottalization

- Uh yeah, yeah, well, it- it- that’s right. And it-
Why fragments are important

- Frequent enough to be a problem:
  - Only 1% of words/3% of sentences
  - But if miss fragment, tend to get surrounding words wrong (word segmentation error).
- Goldwater et al.:
  - 14% absolute increase in word error rate (from 18% to 32%) for words before fragments!!
- Useful for finding other repairs
  - In 40% of SRI-ATIS sentences containing fragments, fragment occurred at right edge of long repair
  - 74% of ATT-ATIS reparanda ended in fragments
- Sometimes are the only cue to repair
  - “leaving at <seven> <fif-> eight thirty”
How fragments are dealt with in current ASR systems

- In training, throw out any sentences with fragments
- In test, get them wrong
- Probably get neighboring words wrong too!
- !!!!!!
Cues for fragment detection

- 49/50 cases examined ended in silence >60msec; average 282ms (Bear et al)
- 24 of 25 vowel-final fragments glottalized (Bear et al)
  - Glottalization: increased time between glottal pulses
- 75% don’t even finish the vowel in first syllable (i.e., speaker stopped after first consonant) (O’Shaughnessy)
Cues for fragment detection

- Nakatani and Hirschberg (1994)
- Word fragments tend to be content words:

<table>
<thead>
<tr>
<th>Lexical Class</th>
<th>Token</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Content</td>
<td>121</td>
<td>42%</td>
</tr>
<tr>
<td>Function</td>
<td>12</td>
<td>4%</td>
</tr>
<tr>
<td>Untranscribed</td>
<td>155</td>
<td>54%</td>
</tr>
</tbody>
</table>
Cues for fragment detection

- Nakatani and Hirschberg (1994)
- 91% are one syllable or less

<table>
<thead>
<tr>
<th>Syllables</th>
<th>Tokens</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>113</td>
<td>39%</td>
</tr>
<tr>
<td>1</td>
<td>149</td>
<td>52%</td>
</tr>
<tr>
<td>2</td>
<td>25</td>
<td>9%</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>0.3%</td>
</tr>
</tbody>
</table>
Cues for fragment detection

- Nakatani and Hirschberg (1994)
- Fricative-initial common; not vowel-initial

<table>
<thead>
<tr>
<th>Class</th>
<th>% words</th>
<th>% frags</th>
<th>% 1-C frags</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stop</td>
<td>23%</td>
<td>23%</td>
<td>11%</td>
</tr>
<tr>
<td>Vowel</td>
<td>25%</td>
<td>13%</td>
<td>0%</td>
</tr>
<tr>
<td>Fric</td>
<td>33%</td>
<td>45%</td>
<td>73%</td>
</tr>
</tbody>
</table>
Liu (2003): Acoustic-Prosodic detection of fragments

- Duration (from alignments)
  Of word, pause, last-rhyme-in word
  Normalized in various ways

- F0 (from pitch tracker)
  Modified to compute stylized speaker-specific contours

- Energy
  Frame-level, modified in various ways
Liu (2003): Acoustic-Prosodic detection of fragments

- **Voice Quality Features**
  - **Jitter**
    - A measure of perturbation in pitch period
    - Praat computes this
  - **Spectral tilt**
    - Overall slope of spectrum
    - Speakers modify this when they stress a word
  - **Open Quotient**
    - Ratio of times in which vocal folds are open to total length of glottal cycle
    - Can be estimated from first and second harmonics
    - Creaky voice (laryngealization) vocal folds held together, so short open quotient
Modulation of vocal fold vibration

- Vocal folds are moved (adducted) by muscles
- Can be tensed – the shorter the vocal folds the faster they vibrate

Slide from Ulrike Gut
Modes of phonation

voicelessness = no vocal fold vibration
modal (normal) voicing
whisper
breathy voice
voice
creaky voice
Breathy voice

- arytenoid cartilages
- remain slightly apart
- continuous airflow
- during vocal fold
- vibration
Creaky voice

- arytenoid cartilages tightly together so that vocal folds can only vibrate at the other end
- irregular, low-frequency vibration

- creaky voice:  

Slide adapted from Ulrike Gut
Liu (2003)

- Use Switchboard 80%/20%
- Downsampld to 50% frags, 50% words
- Generated forced alignments with gold transcripts
- Extract prosodic and voice quality features
- Train decision tree
- Results
  - Precision 74.3%, Recall 70.1%
Liu (2003) features

Features most queried by DT

<table>
<thead>
<tr>
<th>Feature</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>jitter</td>
<td>.272</td>
</tr>
<tr>
<td>Energy slope difference between current and following word</td>
<td>.241</td>
</tr>
<tr>
<td>Ratio between F0 before and after boundary</td>
<td>.238</td>
</tr>
<tr>
<td>Average OQ</td>
<td>.147</td>
</tr>
<tr>
<td>Position of current turn</td>
<td>0.084</td>
</tr>
<tr>
<td>Pause duration</td>
<td>0.018</td>
</tr>
</tbody>
</table>
Liu (2003) conclusion

- Very preliminary work
- Fragment detection is good problem that is understudied!
Fragments in other languages

- Mandarin (Chu, Sung, Zhao, Jurafsky 2006)
- Fragments cause similar errors as in English:

Substitution: 你 - 你 下次 跟他说
you-you next time to him tell
Recognizer output: 那 你 下次 跟他说
that you next time to him tell

- 但我 - 我問的是
- I- I was asking...
- 他是 - 卻很 - 活的很好
- He very- lived very well
Fragments in Mandarin

- Mandarin fragments unlike English; no glottalization.
- Instead: Mostly (unglottalized) repetitions
  
a. 我 - 我问的是 有什么影响
     I - I ask the copula have what influence
     ‘I asked what influence it has.’

b. 他却很好
    he but very well
    ‘But he lives very well.’

- So: best features are lexical, rather than voice quality
Outline

Disfluencies
Characteristics of disfluencies
Detecting disfluencies
Fragments