Language change over the course of membership in an online support group: a corpus linguistic perspective

Daniel McDonald

@interro_gator

This slideshow is available at: http://git.io/v8HnI

11th November 2015
Purpose of today’s talk

1. Introduce myself and my PhD research
   - Literature: online support groups
   - Methods: computational/corpus linguistics
   - Theory of language: systemic functional linguistics
   - Key findings

2. Demonstrate corpkit, a corpus linguistic tool I’ve made

This slideshow is available at: http://git.io/v8HnI
My thesis: case study

- I’m researching **longitudinal linguistic change** in a 12 year old bipolar disorder forum
- 60,000 posts
- 5821 unique usernames
- A few veterans with over 1000 posts, most users have 1–5
- Rules: no asking for diagnosis, stay ‘anonymous’
- Normative biomedical ideology (c.f. Vayreda & Antaki, 2009)
- People come for information and/or social support
- Veterans answer questions, welcome new members, speak as ‘we’
### Bipolar Disorder Message Board

#### HealthBoards > Mental Health Board > Bipolar Disorder

<table>
<thead>
<tr>
<th>Thread / Thread Starter</th>
<th>Rating</th>
<th>Last Post</th>
<th>Replies</th>
<th>Views</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sticky: For Those Recently Diagnosed</strong> amybp</td>
<td><img src="https://via.placeholder.com/15" alt="Image" /></td>
<td>08-30-2013 04:50 AM by amybp</td>
<td>0</td>
<td>1,571</td>
</tr>
<tr>
<td><strong>Sticky: NEW BOARD: Family &amp; Friends of Mentally Ill</strong> Administrator</td>
<td><img src="https://via.placeholder.com/15" alt="Image" /></td>
<td>07-12-2010 10:18 PM by Administrator</td>
<td>1</td>
<td>18,480</td>
</tr>
<tr>
<td><strong>Sticky: Help with Medication or Medical Costs</strong> Administrator</td>
<td><img src="https://via.placeholder.com/15" alt="Image" /></td>
<td>07-08-2009 08:26 PM by Administrator</td>
<td>0</td>
<td>10,322</td>
</tr>
<tr>
<td><strong>Sticky: On this board, negative discussion of bipolar people is inappropriate.</strong> Administrator</td>
<td><img src="https://via.placeholder.com/15" alt="Image" /></td>
<td>10-26-2007 03:20 PM by Administrator</td>
<td>0</td>
<td>16,219</td>
</tr>
<tr>
<td><strong>Sticky: Attention Newbies! Please Read: Bipolar Disorder Symptom Primer</strong> reesie</td>
<td><img src="https://via.placeholder.com/15" alt="Image" /></td>
<td>06-06-2005 08:51 AM by reesie</td>
<td>0</td>
<td>38,990</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Bipolar Disorder?</th>
<th>05-31-2015 02:29 AM by jackie821rrt</th>
<th>1</th>
<th>407</th>
</tr>
</thead>
<tbody>
<tr>
<td>totally screwed up my life</td>
<td>05-30-2015 01:41 PM by subeau</td>
<td>2</td>
<td>479</td>
</tr>
</tbody>
</table>
An example: *Katlin09*

Katlin09’s early posts look like this:
Katlin09

... and here’s one of her later posts:

Re: Psychiatrist vs DR. Psychologist regarding BP Diagnosis

HI. Sorry you’re having such a tough time.

It's not completely unheard of for Pdocs (psychiatrists) to come up with different DX's for the same patient. When you start looking at the DX's you notice they're not all that different.

Your Pdoc DX'd you Bipolar Disorder and Major Depressive Disorder

Bipolar Disorder and Major Depressive Disorder are the same thing. MDD is the Unipolar Depression if your took the Mania away from Bipolar Disorder. When you have Mania and Depression you get Bipolar Disorder. Nowadays this is treated with all types of meds antidepressants, ant-seizure meds, anti-anxiety meds, mood disorder meds, anti-psychotics. The type of med used to treat Bipolar Disorder is not as important as the Cocktail of different meds and what works.
Number of users by number of posts

User’s total post count

Number of users
Total number of posts by membership stage

Total posts by year and membership stage

Number of posts

Year


D. McDonald
UCREL, Lancaster
11th November 2015
The project is situated within four main areas:

- Online communities/support groups/computer-mediated communication (CMC)
- Healthcare communication
- Systemic functional linguistics as theory of language
- Corpus/computational linguistics for tools/methods

There is some research combining some of these areas, but never all four.
Online support groups

There’s been a lot of research into online communities:

- Worries that bad advice is being given by untrained people
- Many studies showing the opposite: ‘go and see your doctor’ (Smithson et al., 2011)
- Unsolicited advice (Vayreda & Antaki, 2009)
- Diverse topics: ‘pro-ana’ (Mantella, 2007), suicide (Horne & Wiggins, 2009)
- Communities of practice, conversation analysis (Stommel, 2008)
- ... and of course, Elena Semino’s current work ...
Why SFL?

SFL attempts to map lexicogrammar to discourse, semantics, register and genre (see Halliday & Matthiessen, 2004). It’s a great approach for this project:

- Handles CMC well: *context is in text* (Eggins, 2004)
- Handles large datasets well: *lexis as delicate grammar* (Hasan, 1987)
- Handles online support groups well:
  - Social support via mood choices
  - Health information provision via transitivity choices
- ‘Plugins’ for multimodality and genre, too!
Harvey has (2007; 2012) investigated a corpus of emails from British adolescents to an online doctor.

- Focus on mental health, sexual health …
- Ascriptions and attributions of depression (2012):
  - *I am depressed*
  - *I have/feel/suffer from depression*

He tends to use simple (though effective!) corpus methods:

- One large dataset
- Reference corpora to determine keywords
- No lemmatisation, annotation parsing, etc.
SFL and corpora

SFL and corpora theoretically work together really well. As explained by Thompson and Hunston (2014), common ground includes:

- Concern with patterns/frequency
- Probabilistic grammar (Halliday, 1991)
- Conceptualisation of register and genre (explicit in SFL, implicit in CL)
- Criticised together by Widdowson (2000) as ‘linguistics applied’

The main things preventing more corpus SFL:

- Automatic parsing of SFL isn’t quite there yet
- Lexical focus of CL
- Collocation and n-grams as typical techniques
- A lot to learn, from opposite sides of campus
Patient, relationship-centred care Matthiessen (2013); Slade et al. (2008)

Mapping landscape of healthcare interactions
- Centred on face-to-face, professional-consumer
- Data collection is ‘expensive’
- So far lacking an account of online/patient–patient interactions

Do corpus linguistics and CMC offer a solution?
Why programming?

Programmatic approaches to research offer:

- **Scalability**: reuse methods on datasets of any size with no additional work
- **Reproducibility**: ensure that anyone, anywhere (including you!) can replicate/modify your results
- **Consistency**: things will be done the same way every time
- **Digital**: as more and more datasets are ‘born digital’, programming becomes more useful

Python in particular is great for:

- Being language-like, modern, readable
- Working with texts (*NLTK*)
- Working with datasets (*pandas*)
- Visualisation (*matplotlib*)

D. McDonald
UCREL, Lancaster
11th November 2015
Why make tools?

There are a lot of tools already out there. Why make more?

- Few tools can:
  - Work with parsed data
  - Work with subcorpora
  - Translate annotations to functional grammatical concepts

- More links needed to high-level modules for manipulating and visualising data

- Added flexibility of a full programmatic interface

- The importance of free, open-source tools, version control, etc.
Case study: the Bipolar Forum

- Every post to the forum was annotated with grammatical structure:

  ![Diagram of annotated sentence]

- Sorted posts into ten subcorpora, representing ten stages of membership
- Sorted threads by date, speaker segmented, in order to look at evolution of norms
- corpkit was created to work with the data:
  - https://www.github.com/interrogator/corpkit
  - http://interrogator.github.io/corpkit/
Interpersonal meanings: modalisation

Modals when subject is I

- could/can
- would/will
- might/may/must
- should/shall

Percentage of all modals

Group
Socio-semiotic activities in

I would + adjunct

Request for information:

1. I would very much appreciate any insight you can give us.
2. I would sure appreciate any information on the disability question.
3. I guess I would just like to hear if anyone has been where I’m at.

Reporting, sharing:

1. I would verbally crush anyone who would... Speak.
2. In the middle of my explosions, I would always ask internally, who is this person.
3. I would actually leave my house and socialize with people, which I hadn’t done comfortably in years.

Expounding, recommending:

1. I would DEFINITELY recommend seeing a psychologist.
2. I would certainly make a point to follow up.
3. I would definitely have your daughter pay her
Experiential meanings: being and having bipolar disorder

Intensive attribution → Possessive attribution
# import my module
from corpkit import *
from dictionaries.process_types import processes as p

# make a query
bp_words = r'(?i)\b(bipolar|bi-polar|bp|b-p)'

# search the corpus
govs = interrogator(corpus, 'g', bp_words, lemmatise = True)

# edit results
rename = [(r'root:root', 'be bipolar'),
          (r'(dobj|acomp):have', 'have bipolar'),
          (r'dobj:%s' % p.relational, 'other relational processes')]

be_have = editor(govs.results, '%', 'self', sort_by = 'total',
                 replace_names = rename, just_entries = [n for r, n in rename])

# visualise
plotter('Being and having bipolar', be_have.results,
        y_label = 'Percentage of all relational processes')
Evolution of *having bipolar*
Proportion of posts by membership stage

Total posts by year and membership stage

Year

Percentage


0 10 20 30 40 50 60 70 80 90 100 110 120

Percentage

Year


D. McDonald
UCREL, Lancaster
11th November 2015
Hi, I’m new to this site also.
I was diagnosed Bi polar 17 years ago after a suicide attempt
and have bounced around between denial and treatment for a long time.
I’ve been under strict dr.’s care and med. treatment for the last 6
months, because the depression has been so bad at times I can’t leave
my bed.
I’m 38, with 2 kids and not much of a support system at home.
I looked up this site, looking for someone to talk to that might
understand what I was going through.
I’m currently on the medicine merry go round,
just came off of Lamictal, now on Serroquel, Zoloft, Topomax, and
Ambien.
I also have several other chronic medical conditions that
I have to take meds daily for which include Fentanyl patches and pain
meds for chronic pain...
these things make the depression worse.
Identifying key Participants

<table>
<thead>
<tr>
<th>Post 1</th>
<th>L/L</th>
<th>Post 560+</th>
<th>L/L</th>
</tr>
</thead>
<tbody>
<tr>
<td>i</td>
<td>1972.09</td>
<td>you</td>
<td>4179.76</td>
</tr>
<tr>
<td>anyone</td>
<td>326.47</td>
<td>pdoc</td>
<td>1584.74</td>
</tr>
<tr>
<td>he</td>
<td>224.63</td>
<td>tdoc</td>
<td>718.10</td>
</tr>
<tr>
<td>she</td>
<td>113.17</td>
<td>kat</td>
<td>655.99</td>
</tr>
<tr>
<td>mother</td>
<td>102.18</td>
<td>kait</td>
<td>633.67</td>
</tr>
<tr>
<td>medication</td>
<td>97.49</td>
<td>erin</td>
<td>566.55</td>
</tr>
<tr>
<td>myself</td>
<td>95.50</td>
<td>she</td>
<td>444.97</td>
</tr>
<tr>
<td>doctor</td>
<td>89.19</td>
<td>hug</td>
<td>412.99</td>
</tr>
<tr>
<td>swing</td>
<td>81.92</td>
<td>pdocs</td>
<td>355.54</td>
</tr>
<tr>
<td>bipolar</td>
<td>64.84</td>
<td>med</td>
<td>339.77</td>
</tr>
<tr>
<td>episode</td>
<td>59.19</td>
<td>bp</td>
<td>305.23</td>
</tr>
<tr>
<td>this</td>
<td>54.82</td>
<td>goody</td>
<td>292.57</td>
</tr>
<tr>
<td>disorder</td>
<td>54.66</td>
<td>daughter</td>
<td>273.44</td>
</tr>
<tr>
<td>psychiatrist</td>
<td>52.05</td>
<td>able</td>
<td>198.95</td>
</tr>
<tr>
<td>problem</td>
<td>50.13</td>
<td>stability</td>
<td>186.17</td>
</tr>
</tbody>
</table>
Common participants in veteran posts

Percentage of all words

- **we**: 0.37
- **thing(s)**: 0.24
- **pdoc(s)**: 0.21
- **bp(s)**: 0.09
- **us**: 0.08
- **tdoc(s)**: 0.04
- **manic/mania(s)**: 0.02
- **board**: 0.01
- **hypomanic/hypomania(s)**: 0.01
...and key Events/Processes

<table>
<thead>
<tr>
<th>Post 1</th>
<th>L/L</th>
<th>Post 560+</th>
<th>L/L</th>
</tr>
</thead>
<tbody>
<tr>
<td>diagnose</td>
<td>763.75</td>
<td>use</td>
<td>77.40</td>
</tr>
<tr>
<td>appreciate</td>
<td>166.82</td>
<td>call</td>
<td>44.89</td>
</tr>
<tr>
<td>suffer</td>
<td>130.24</td>
<td>suggest</td>
<td>42.44</td>
</tr>
<tr>
<td>be</td>
<td>101.43</td>
<td>do</td>
<td>41.15</td>
</tr>
<tr>
<td>wonder</td>
<td>63.79</td>
<td>allow</td>
<td>35.87</td>
</tr>
<tr>
<td>start</td>
<td>39.46</td>
<td>stabilize</td>
<td>31.64</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>hope</td>
<td>-39.69</td>
<td>hide</td>
<td>-19.51</td>
</tr>
<tr>
<td>say</td>
<td>-41.47</td>
<td>suffer</td>
<td>-23.29</td>
</tr>
<tr>
<td>see</td>
<td>-46.20</td>
<td>diagnose</td>
<td>-33.30</td>
</tr>
<tr>
<td>help</td>
<td>-49.26</td>
<td>love</td>
<td>-43.41</td>
</tr>
<tr>
<td>keep</td>
<td>-51.38</td>
<td>lose</td>
<td>-58.35</td>
</tr>
<tr>
<td>do</td>
<td>-69.38</td>
<td>try</td>
<td>-67.28</td>
</tr>
</tbody>
</table>
Diagnose as process

1. My wife was diagnosed as being bipolar back in December when I had to have her hospitalized for severe depression.
2. I was diagnosed with Disthymia at that time.
3. My 10 year old has been diagnosed with BP and depression.
4. My son was diagnosed with bipolar disorder and anxiety disorder today, they gave me a script for risperidone, but the side effects appear to be pretty scary.
5. I was diagnosed with bipolar disorder at 20 years old.
6. My father was diagnosed with BPII at the age of 31, when I was 1 years old.
What circumstances co-occur with diagnose?

Circumstances surrounding the process of diagnosis

- First posts
- Veteran posts (groups 6+)

Percentage of all diagnose circumstances vs. Word
Thesis: key findings

- Shift from information seeking to social support providing:
  - Socio-semiotic activities: *Sharing & Reporting* → *Expounding & Recommending*
- Jargonisation, hedged commands, metadiscourse, vague language
- *Being/suffering* → *having* bipolar
- *Mood swings/episodes* → *stabilisation and balance*
- Maintaining a complex ‘lay-expert’ identity
- Fostering empowerment over health problems
- Focus on actions in the present, positive mental processes
Stuff I don’t have time to explore in more detail

- Parser accuracy
- Systemic parser
- Genre
- Multimodality
- Appraisal
- Rank shift/grammatical metaphor:
My really lovely pdoc helped me out.

My pdoc is really lovely and helped me out.

My pdoc, being really lovely, helped me out.

My pdoc helped me out. She’s really lovely.

With each example, there is not only a need for more complex kinds of corpus interrogation, but also an increased likelihood of parsing errors that may prevent automatic location of the string.
Over to GitHub...

1. https://www.github.com/interrogator/corpkit
Links

Contact:

- Email: mcdonaldd@unimelb.edu.au
- Twitter: @interro_gator
- GitHub: https://www.github.com/interrogator
- Academia.edu: https://unimelb.academia.edu/danielmcdonald

Work:

- corpkit: http://interrogator.github.io/corpkit/
- Slides: http://git.io/v8HnI
- risk: http://goo.gl/NsH1UK


References IV


