

Høgskolen i Hedmark



Weighing ethical considerations in corpus research based on online data sources

Susan Nacey

BAAL; September 2016

Internet research

1. Research on the phenomenon of the internet
 - *Exceptional new object*
2. Use of the internet as a tool in conducting research
 - *Instrumentation*
3. Research on production via the internet
 - *Object of study*



The language of betrayal

- 47.5 million words
- 4845 individual informants
- 6-year period



Research questions

- What metaphorical expressions do survivors of relationship abuse use when communicating about their experience?
- What may such metaphorical expressions reveal about the attitudes of relationship abuse survivors towards e.g.:
 - themselves
 - their abusers
 - third parties
 - their experience of abuse
 - their recovery
- How may researchers harness 'big data' to study metaphor?
 - Technically
 - Ethically

Internet research ethics

- Guidelines:

- institutional review boards (IRBs)
- national agencies
- academic societies
- individual researchers

- Norway:

- [The Norwegian National Research Ethics Committees](#)

National Committee for Research Ethics in the Social Sciences, Humanities; Law and Theology (NESH)

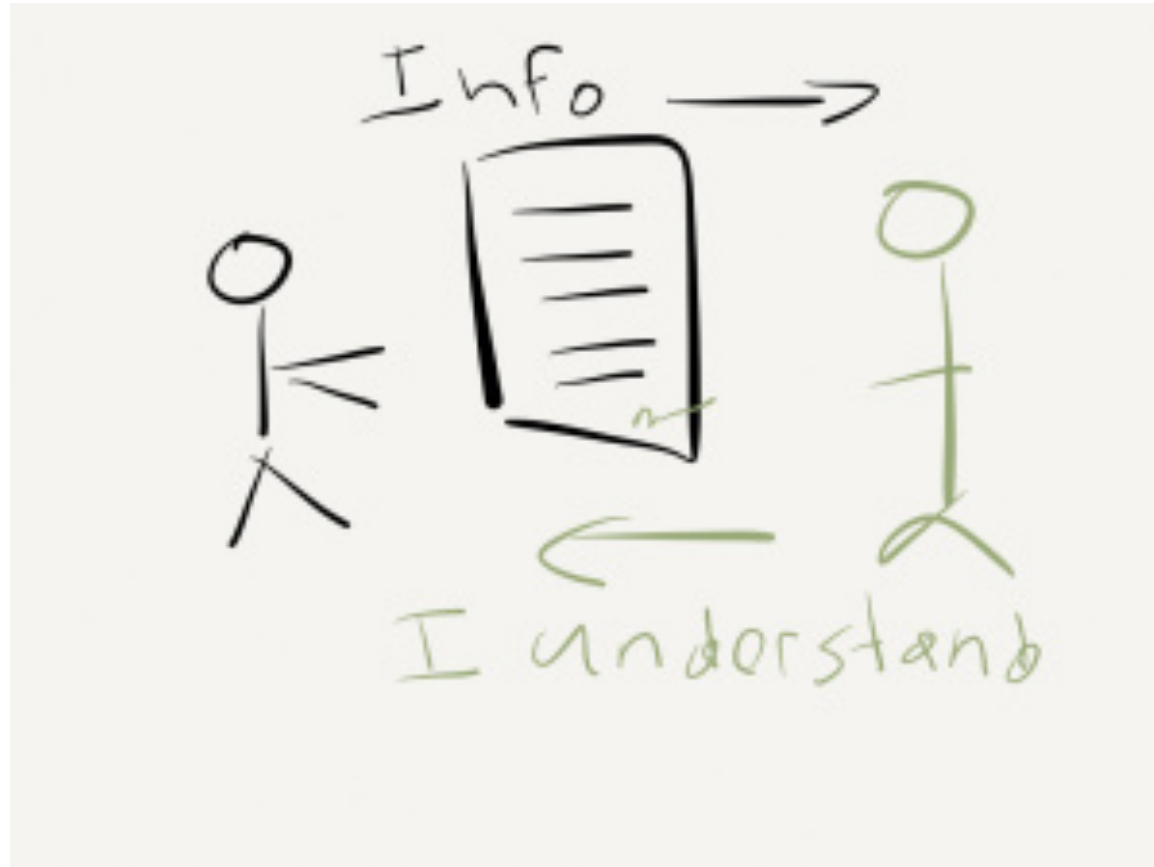
- The law

[The Personal Data Act](#)

Weighing different considerations...



Informed consent



Sensitive topics



Public vs. Private

- Perceived privacy
- Accessibility
- Private places in a public forum



Terms and conditions

40%

of consumers agree that they never read any terms and conditions before signing up to online services

“These terms of service aren’t laws, they’re statements written by Web site owners describing what they’d like to happen if they ran the universe.”

Professor Christian Sandvig,
University of Michigan

From: “Why I am suing the government (or: *I write scripts, bots, and scrapers that collect online data*)” (01 July 2016)

Anonymization



Levels of disguising

- None
- Light
- Moderate
- Complete

Bruckman 2002

Alternatives?

- Establish a 'research' forum
 - The Observer's Paradox
- Disguise completely
- Reconsider dissemination channels
- Abandon the research



What animals/living creatures do survivors of relationship abuse draw upon to communicate about their experience?

ALBATROSS (3) **ANIMAL** (84) ANT (7) **BEAR** (40) **BEAST** (73) BEE (3) BIRD (30) BULL (39) BUNNY (19)
BUTTERFLY (22) CAMEL (17) **CAT** (101) CATFISH (1) CHAMELEON (24) **CHICK** (52) CHICKEN (21)
COBRA (3) COCKROACH (28) COW (35) **CREATURE** (188) CRITTER (1) CROCODILE (5) CUCKOO (2)
DEER (12) **DOG** (126) DOLPHIN (2) DONKEY (4) DOVE (4) **DUCK** (47) DUCKLING (6) ELEPHANT (22) **FISH** (73)
FLOUNDER (1) **FLY** (42) FOX (4) **FROG** (47) GAZELLE (6) GOOSE (2) GUINEA PIG (5) HAMSTER (11) HAWK (8) HEN (2)
HORSE (44) HYENA (7) KITTEN (2) LAMB (15) **LEECH** (39) LION (31) LIZARD (6) MARE (1) **MONKEY** (62)
MONSTER (1245) MOSQUITO (6) MOTH (9) **MOUSE** (80) OWL (2) OYSTER (5)
PARASITE (134) PARROT (3) PEACOCK (6) PENGUIN (3) PET (7) **PIG** (82) PITBULL (5) PONY (3) POODLE (5)
PREDATOR (501) **PREY** (263) PUPPY DOG (7) **PUPPY** (49) PUSSY (6) RABBIT (6) **RAT** (76)
SCORPION (13) SERPENT (5) **SHARK** (63) **SHEEP** (45) SHREW (4) SLUG (6) SNAIL (12) **SNAKE** (102)
SPIDER (13) SQUIRREL (2) **SWINE** (10) TIGER (6) TITMOUSE (2) TOAD (13) TURTLE (3) **VULTURE** (28) WEASEL (3) WHALE (5)
WOLF (82) **WORM** (18)



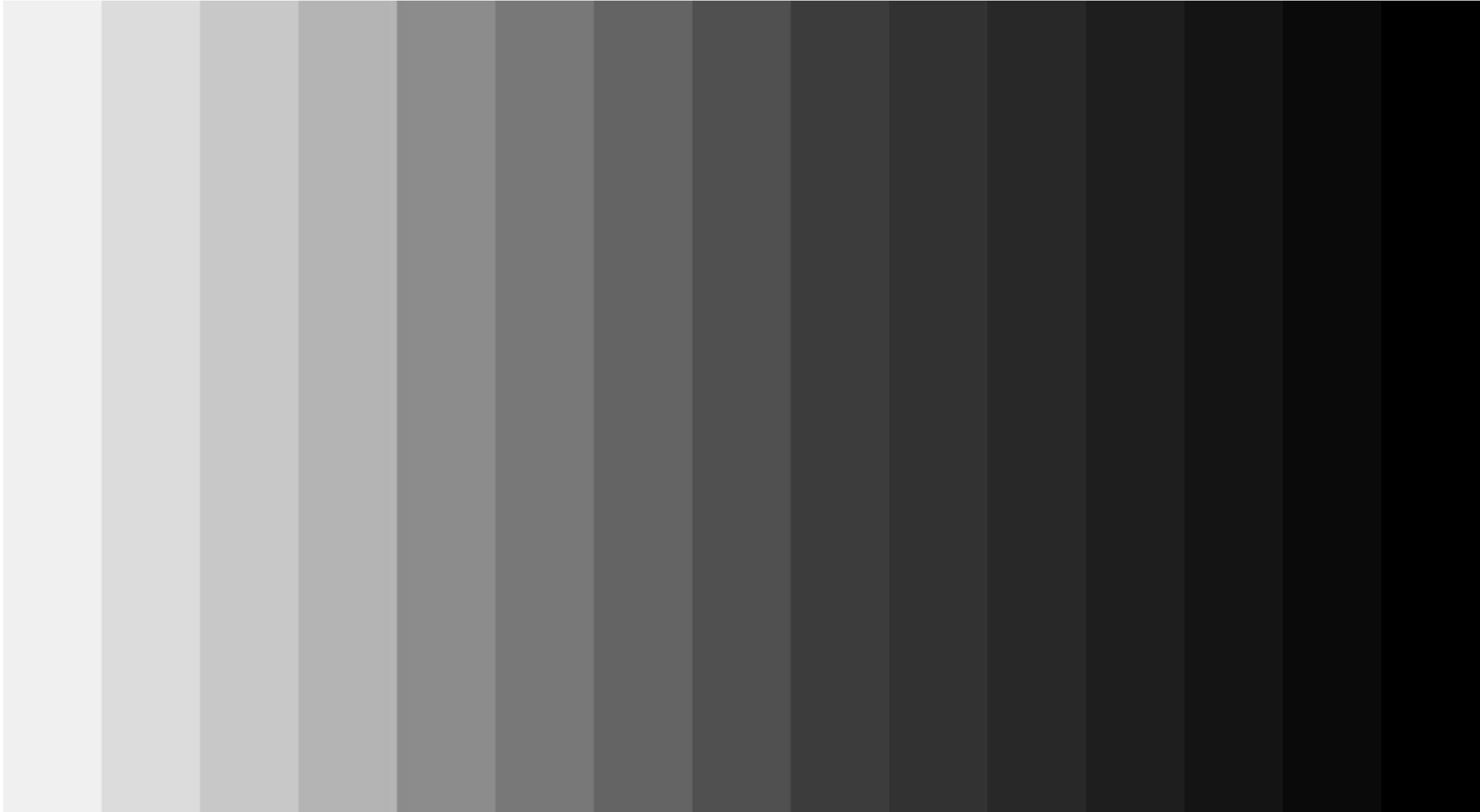
Scientific value



Utility to society



Concluding thoughts



Selected references

- Bruckman, A. (2002). Ethical Guidelines for Research Online. Retrieved from <http://www.cc.gatech.edu/~asb/ethics/>
- Halvarson, C., & Lilliengren, P. (2003). Private explanatory systems and informed consent online. An ethical and methodological discussion. In M. Thorseth (Ed.), *Applied Ethics in Internet Research* (pp. 125-139). Trondheim: Norwegian University of Science and Technology.
- Mann, C. (2003). Generating data online. Ethical concerns and challenges for the C21 researcher. In M. Thorseth (Ed.), *Applied Ethics in Internet Research* (pp. 31-49). Trondheim: Norwegian University of Science and Technology.
- Herring, S. (1996). Linguistic and critical analysis of computer-mediated communication: Some ethical and scholarly considerations. *The Information Society*, 12, 153-168.
- NESH - The National Committee for Research Ethics in the Social Science and the Humanities. (2016). *Guidelines for Research Ethics in the Social Sciences, Humanities, Law and Theology* Retrieved from https://www.etikkom.no/globalassets/documents/english-publications/60127_fek_guidelines_nesh_digital_corr.pdf Retrieved from https://www.etikkom.no/globalassets/documents/english-publications/60127_fek_guidelines_nesh_digital_corr.pdf
- Sandvig, C. (2016). Why I am Suing the Government (or: I write scripts, bots, and scrapers that collect data online). Retrieved from <https://socialmediacollective.org/2016/07/01/why-i-am-suing-the-government/>
- Segadal, K. U. (2015). Possibilities and limitations of Internet research: A legal framework. In H. Fossheim & H. Ingierd (Eds.), *Internet Research Ethics* (pp. 35-47). Oslo: Cappelen Damm Akademisk. Retrieved from <http://press.nordicopenaccess.no/index.php/noasp/catalog/book/3>. Retrieved from <http://press.nordicopenaccess.no/index.php/noasp/catalog/book/3>

Material and methods

Wmatrix3

ucel.lancs.ac.uk/wmatrix3.html

Tell_Your_Story > Viewing Word-SEM frequency profile

Wmatrix

You are logged in to Wmatrix3 as: hihmney

[Tagging > Tag Wizard... | My Tag Wizard... | Domain Tag Wizard... | Load file without tagging...]
 [Folders > My folders | Details | Create... | Delete... | Archive... | Extract... | Join... | Share... | NEW! CrossTab... | Empty TRASH]
 [Options > Switch to Simple Interface | Edit user options...]
 [Help > Contents | Availability | Tagsets: POS & Semantic | USAS: Lexicon & MWEs & Context rules | Updates | Feedback]

[You are here > My folders > Tell_Your_Story | File details]

Save

Search term: '.L2'.
Sorted on USAS tag.

You are viewing a frequency profile.
Click on a column heading to sort on that column.
Click on a Concordance link to see concordance lines.
Please note that concordances are not filtered by tags, so will contain all occurrences of the word.

BROAD SWEEP
(find tags anywhere on a word's tag-list)

Search shortcuts
(searches for top-level semantic categories):
Show complete list

Broad Sweep Search:
Enter the tag you wish to search for here:
Go
(you can also search for part of a tag)

Search shortcuts:
Show complete list

Search this list:
Enter the word or tag you wish to search for here:
L2 Go
(you can also search for part of a word or tag; enter '.' or leave blank for complete list)

Remember your last search:
To remember the search currently shown on the right, give it a name in the box below and press Go.
The search will be saved in the advanced folder view:
L2 Go

You can use regular expressions in the search box.
Help on regular expressions is available at many websites, e.g. [regular-expression.info](#)

Word	Sentag	Frequency	Relative Frequency	Concordance
Asp	L2	1	0.00	Concordance
Boston_terrier	L2	1	0.00	Concordance
Jack_Russell	L2	1	0.00	Concordance
MRSs	L2	1	0.00	Concordance
Maine_Coon	L2	1	0.00	Concordance
Scorsese	L2	1	0.00	Concordance
Tabby	L2	1	0.00	Concordance
Yentas	L2	1	0.00	Concordance
Yorkshire_terrier	L2	1	0.00	Concordance
amoeba	L2	1	0.00	Concordance
anted	L2	1	0.00	Concordance
antelopes	L2	1	0.00	Concordance
ape	L2	1	0.00	Concordance
apes	L2	1	0.00	Concordance
bacterias	L2	1	0.00	Concordance
basset	L2	1	0.00	Concordance
bilirubin	L2	1	0.00	Concordance
birder	L2	1	0.00	Concordance
birdies	L2	1	0.00	Concordance
birdwatch	L2	1	0.00	Concordance
bloodhounds	L2	1	0.00	Concordance
bloodsuckers	L2	1	0.00	Concordance
boa	L2	1	0.00	Concordance
boar	L2	1	0.00	Concordance
bottlenose	L2	1	0.00	Concordance
bray	L2	1	0.00	Concordance
bristles	L2	1	0.00	Concordance
bugdies	L2	1	0.00	Concordance
bunny_rabbit	L2	1	0.00	Concordance
buzzard	L2	1	0.00	Concordance
calves	L2	1	0.00	Concordance
canary	L2	1	0.00	Concordance
carcasses	L2	1	0.00	Concordance
carcasses	L2	1	0.00	Concordance
carp	L2	1	0.00	Concordance
cat_pictures	L2	1	0.00	Concordance
cat_sit	L2	1	0.00	Concordance
cat_sitting	L2	1	0.00	Concordance
cattery	L2	1	0.00	Concordance
catting	L2	1	0.00	Concordance
chameleons(p)	L2	1	0.00	Concordance
chick	L2	1	0.00	Concordance
chickadee	L2	1	0.00	Concordance

Summary information:

Number of types shown: 602
 Total frequency of types shown: 11782 (0.12%)
 Total frequency overall: 9636434

Number of items shown with a given frequency:

Frequency	Types	Tokens
1	215 (35.71%)	215 (1.82%)
2	91 (15.12%)	182 (1.54%)
3	41 (6.81%)	123 (1.04%)
4	37 (6.15%)	148 (1.26%)
5	30 (4.98%)	150 (1.27%)
6	22 (3.65%)	132 (1.12%)
7	14 (2.33%)	98 (0.83%)
8	10 (1.66%)	80 (0.68%)
9	11 (1.83%)	99 (0.84%)
10	8 (1.33%)	80 (0.68%)
> 10	123 (20.43%)	10475 (88.91%)

76 Identify Post

Original text file: C:/Users/susanlm/Dropbox/Betrays Choose file

Concordance xlsx file: C:/Users/susanlm/Dropbox/Betrays Choose file

Output csv file: birds.csv Choose file

Process files

```

Loading & preparing original file...
Loading & preparing concordance file...
Searching original file for concordance lines...
Some concordance lines not found.
Searching again with one less word at left and right of each line...
Some concordance lines not found.
Searching again with one less word at left and right of each line...
Writing results to birds.csv...
  
```

FileMaker Pro Advanced

File Edit View Insert Format Records Scripts Tools Window Help

Language of betrayal

Records 45 / 7798 Found (Unsorted)

Layout: Language of betrayal View As: Preview

Analyzed N Y Folder: TYS Semantic tag: L2 Living creatures: animals, birds, etc.

Left co-text: ords, including her bank accounts (she 's loaded !) Can you say cash

Token: cow

Right co-text: , ATM machine and meal ticket ? Yes, I believe those labels belong to h

LexemeType: Specific (adult female)

Lexeme: COW

Conventionalized idiom/coll: Y Variation Conv metonymy

Metaphor: Y N O DK Metonymy: Y N Notes: Y N

Topic: Person

TPerson: OW

TSpec: Particular

Time: N/A

MetType: Indirect Direct N/A Both

DirDelMet:

Mflag:

Metallanguage: Y N

Explanation/plural analogies: Yes No Memorable: Yes No

Rejected: N Y Check Links:

TokenType: n

ID: 2700

Perspective: Survivor

PSpec: Particular (poster)

Attitude: Negative

Negation: Y N

Poster: 660

Thread: 17883

Can-you-con-a-psychopath

Popular culture reference: Y N

Question: Y N

Complimentary roles: Y N

SurvivorRole:

AbuserRole:

CultRef:

RStudio File Edit Code View Plots Session Build Debug Tools Window Help

~/Dropbox/Betrayal metaphors/Amets - RStudio

```

315 AS$Neut <- NULL
316 AS$Sum <- NULL
317 AS$NegF <- NULL
318 AS$PosF <- NULL
319 AS$NeutF <- NULL
320 for (i in AS$ID) {
321   AS$ASNeg[AS$ID==i] <- sum(amets$Attitude[amets$PosterID==AS$Poster[AS$ID==i] &
322     amets$Lexeme==AS$Lexeme[AS$ID==i]
323     & amets$TPerson == "Abuser" & amets$Perspectiv
324   AS$ASPos[AS$ID==i] <- sum(amets$Attitude[amets$PosterID==AS$Poster[AS$ID==i] &
325     amets$Lexeme==AS$Lexeme[AS$ID==i]
326     & amets$TPerson == "Abuser" & amets$Perspectiv
327   AS$ASNeut[AS$ID==i] <- sum(amets$Attitude[amets$PosterID==AS$Poster[AS$ID==i] &
328     amets$Lexeme==AS$Lexeme[AS$ID==i]
329     & amets$TPerson == "Abuser" & amets$Perspectiv
330
331   ## Calculates sums and frequencies
332   AS$Sum <- rowSums(AS[,c("ASNeg", "ASNeut", "ASPos")])
333   AS$NegF <- AS$ASNeg / AS$Sum
334

```

Environment History

col.main = "darkgreen", xlab = "Posters", col.lab = "darkgreen")
legend("topleft", legend = "307 posters\n492 observations")
plot(sort(LP0\$SSPosIndex), col="darkgreen", type = "l", lwd = 5,
main = "Topic = Survivor", ylab = "Positivity index",
col.main = "darkgreen", xlab = "Posters", col.lab = "darkgreen")
legend("topleft", legend = "75 posters\n94 observations")

Files Plots Packages Help Viewer

Zoom Export Publish

Topic = Survivor

Positivity index

Posters

75 posters
94 observations

Console

```

~/Dropbox/betrayal metaphors/Amets /
> barplot(SurvPost~SurvObs)
Error in barplot.default(SurvPost ~ SurvObs) :
'height' must be a vector or a matrix
> ##plot Survivor attitude
> plot(sort(LP0$SSNegIndex), col="darkgreen", type = "l", lwd = 5,
+ main = "Topic = Survivor", ylab = "Negativity index",
+ col.main = "darkgreen", xlab = "Posters", col.lab = "darkgreen")
+ legend("topleft", legend = "307 posters\n492 observations")
>
>
> plot(sort(LP0$SSPosIndex), col="darkgreen", type = "l", lwd = 5,
+ main = "Topic = Survivor", ylab = "Positivity index",
+ col.main = "darkgreen", xlab = "Posters", col.lab = "darkgreen")
+ legend("topleft", legend = "75 posters\n94 observations")
>

```

A's abuser = 'Jaws'

- **B:** I started calling the X "Jaws" too. I had a dream about a great white shark circling me off and on during the relationship with him. (The good old subconscious!!) They are very much like sharks, even to the dead look they get in their eyes!!!
- **C:** Maybe, our subconscious picked up on cognitive dissonance during the relationship and was letting us know a predator is out there. I mean, is there a better predator on the face of the earth? Course, my Xs not of the caliber of a great white. Hes more like a nurse shark, or is that a nasty label to those nurse sharks? I mean, what did they ever do to us?
- **A:** You're right. Now I feel bad for putting down sharks. I guess I've always been so afraid of them because I came from the generation when JAWS was in theaters so pop culture was everything JAWS. We'd play JAWS in the pool instead of Marco Polo. You think your ex is a nurse shark, huh? Did you know they're bottom feeders?
- **C:** As to knowing about Nurse Sharks being bottom feeders, um, yeah, you could say I knew...HA!: