Linguistic fingerprints across the whole of life: Analysing the language used in childhood essays and its predictive power for the future

Alissa Goodman, Margaret L Kern, Martina Narayanan, Benedetta Pongiglione, JD Carpentieri, H. Andrew Schwartz

The Transformative Essays Project

CENTRE FOR LONGITUDINAL STUDIES

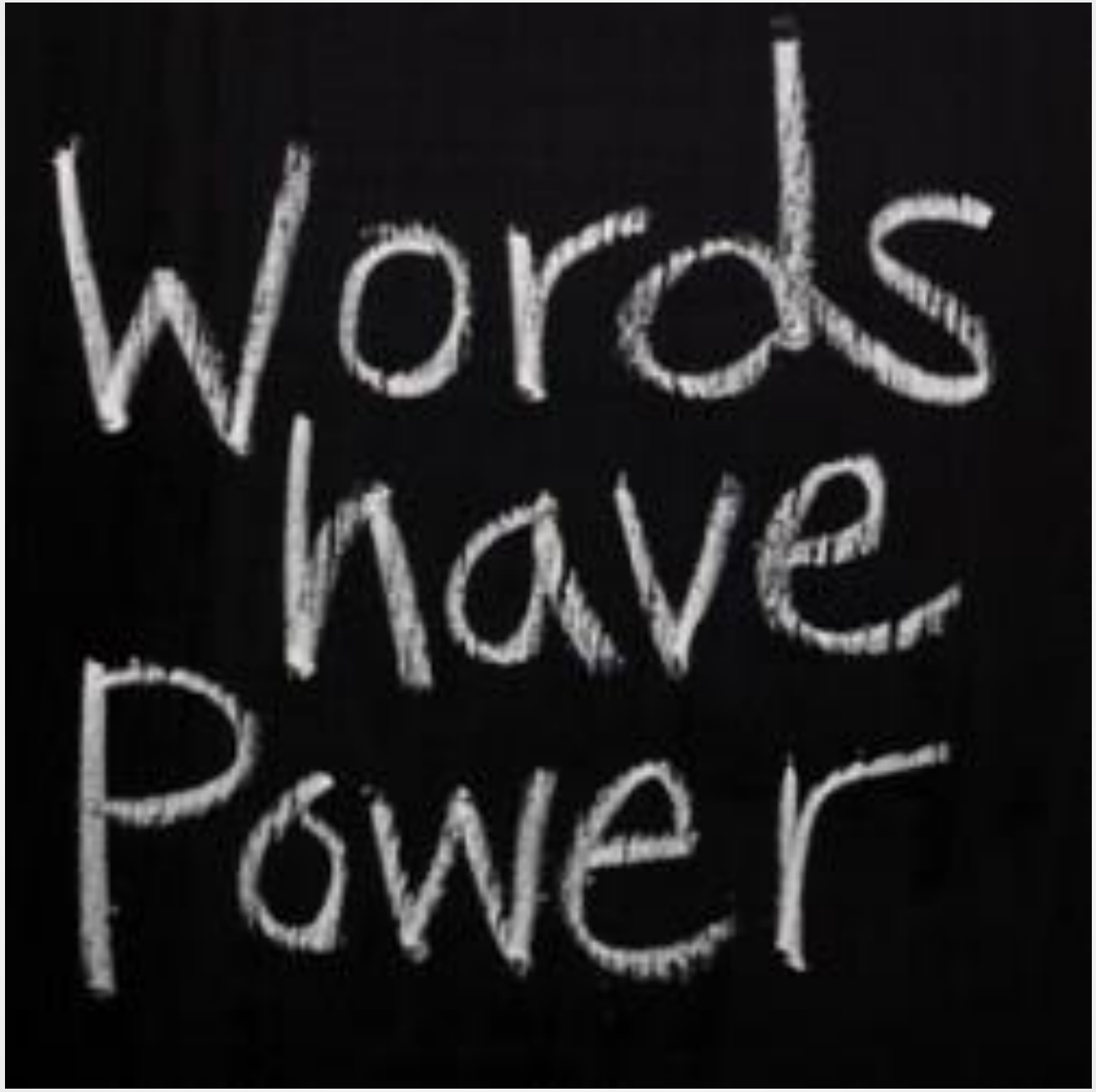
Martina Narayanan, Alissa Goodman, JD Carpentieri, Benedetta Pongiglione

Stony Brook University

H. Andrew Schwartz, Swatilekha Chaudhury, Neelaabh Gupta, Anvesh X

THE UNIVERSITY OF MELBOURNE

Margaret Kern


Words
have
Power


TIME
NICOLE DOES THE CANCAN
BELIEVE IT OR NOT, THIS 91-YEAR-OLD NUN CAN HELP YOU BEAT
ALZHEIMER'S
A landmark study of the disease sheds new light on:
WHAT CAUSES IT
HOW TO PREVENT IT

Analysing School Mission Statements

“Our purpose is to provide a supportive learning environment where young people are empowered to achieve their personal best and develop as responsible and creative members of our community. In achieving our purpose we value: mutual respect, focus on learning, achieving our personal best, responsibility for our actions, and acting with integrity.”

Analysing School Mission Statements

Factor	Coding rubric
A = Academic Motivation	1 = Foster Cognitive Development 2 = Academic Success and Performance 3 = Acquire Knowledge (e.g. learning) 4 = Academic Self-Discipline and Regulation, Motivation, Drive 5 = Reach Potential 6 = Expert Academic Success 7 = Valuing Academics (perceived instrumentality)
B = Personal Characteristics	1 = Positive Affect (e.g. Happiness) 2 = Productive Coping, Problem Solving 3 = Self-Sufficient 4 = Self-Efficacy, Confidence 5 = Self-Esteem 6 = Hope/Hopefulness 7 = Adaptability 8 = Adjustment to School, Flexibility

Allen, Kern, Vella-Brodrick, Waters, 2018

Analysing School Mission Statements

"Our purpose is to provide a supportive (E6) learning (A3) environment (J2) where young people are empowered to achieve their personal best (A5) (O1) and develop as responsible (G4) and creative members of our community (C5). In achieving our purpose we value: mutual respect, focus on learning (A3), achieving our personal best (A5), responsibility for our actions, and acting with integrity (G4)."

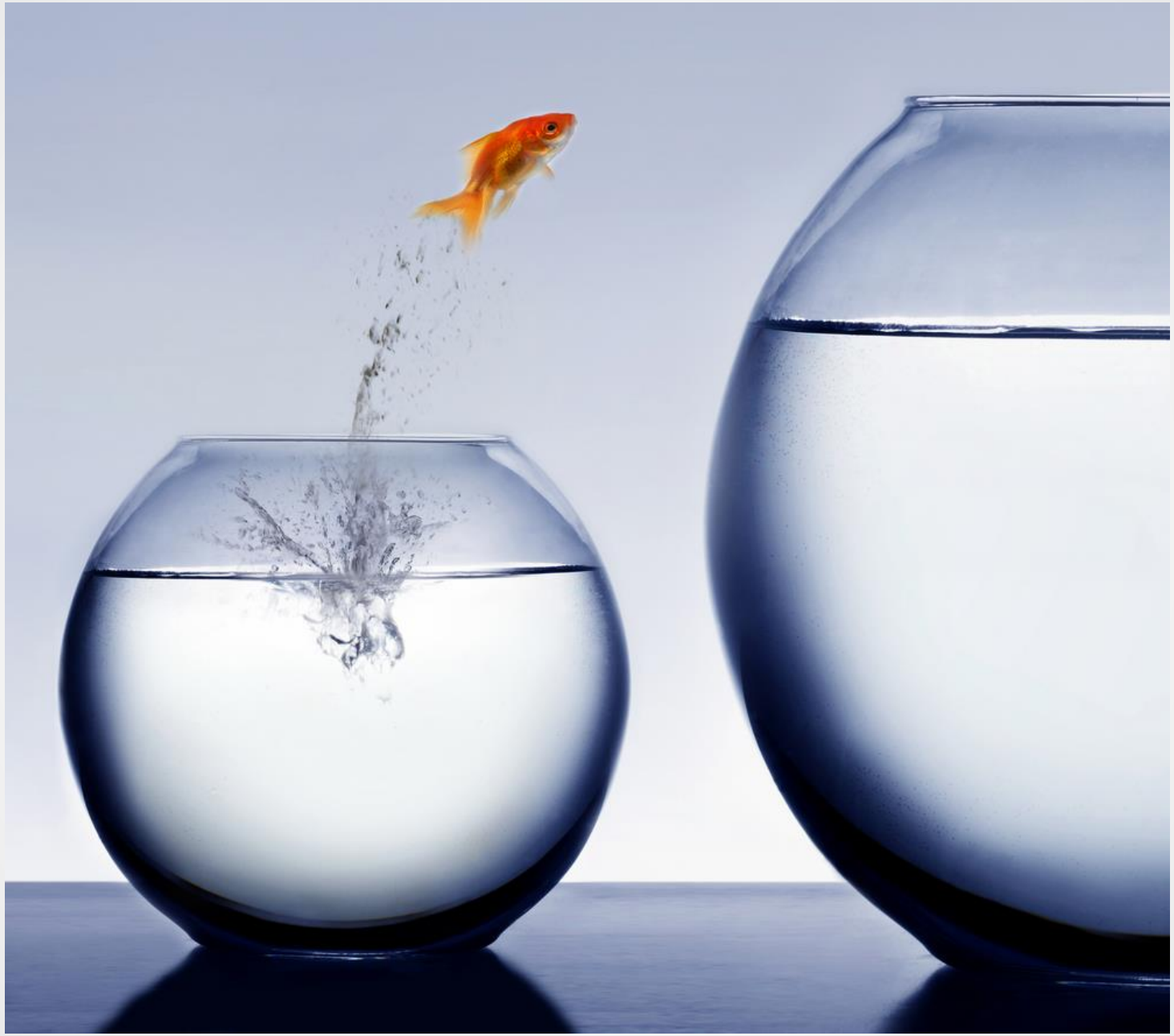
Allen, Kern, Vella-Brodrick, Waters, 2018

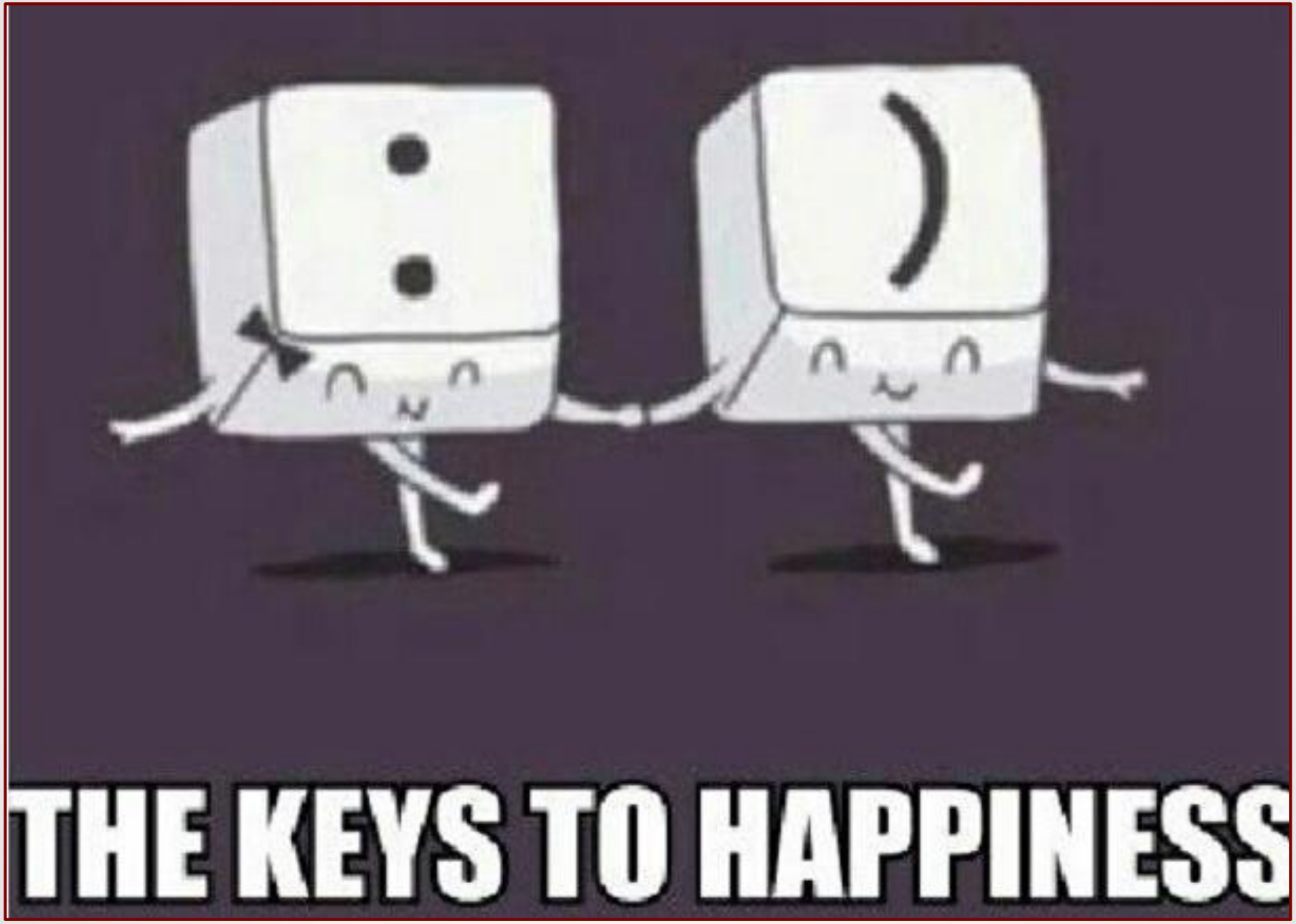
Automating the coding process

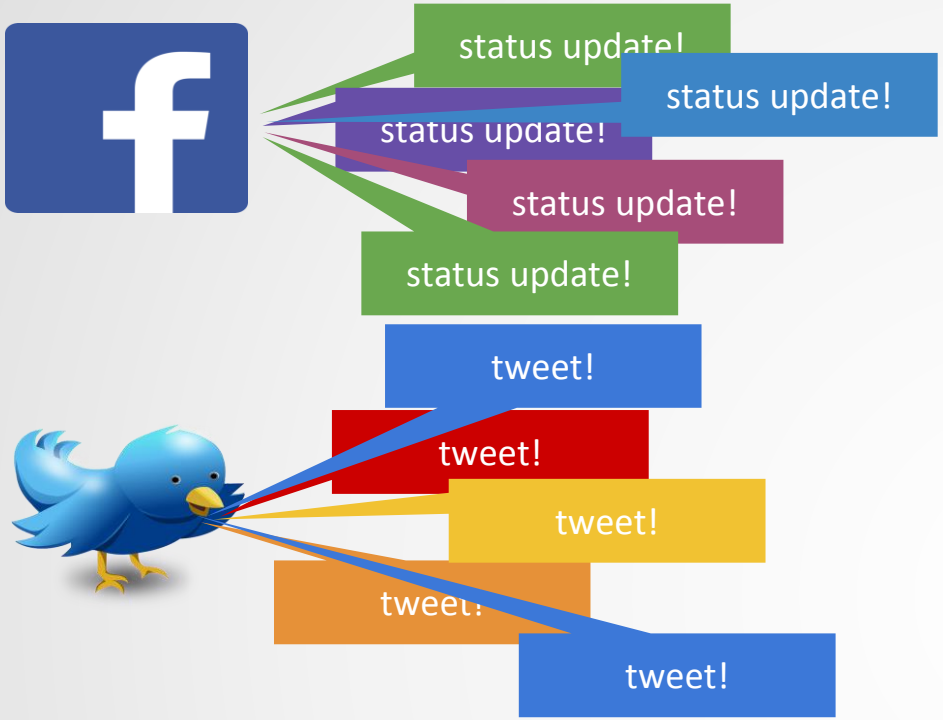
Awesome	Great	Hope	Please
Party	Respect	Comfort	Forgive
Pride	Happy	Super	Likes
Positive	Cares	Strong	Enthusiastic
Good	Satisfied	Humour	Thanks
Thanks	Funny	Trust	Easy
Love	Optimism	Wonderful	Fun
Enjoy	Luck	Yay	Joy

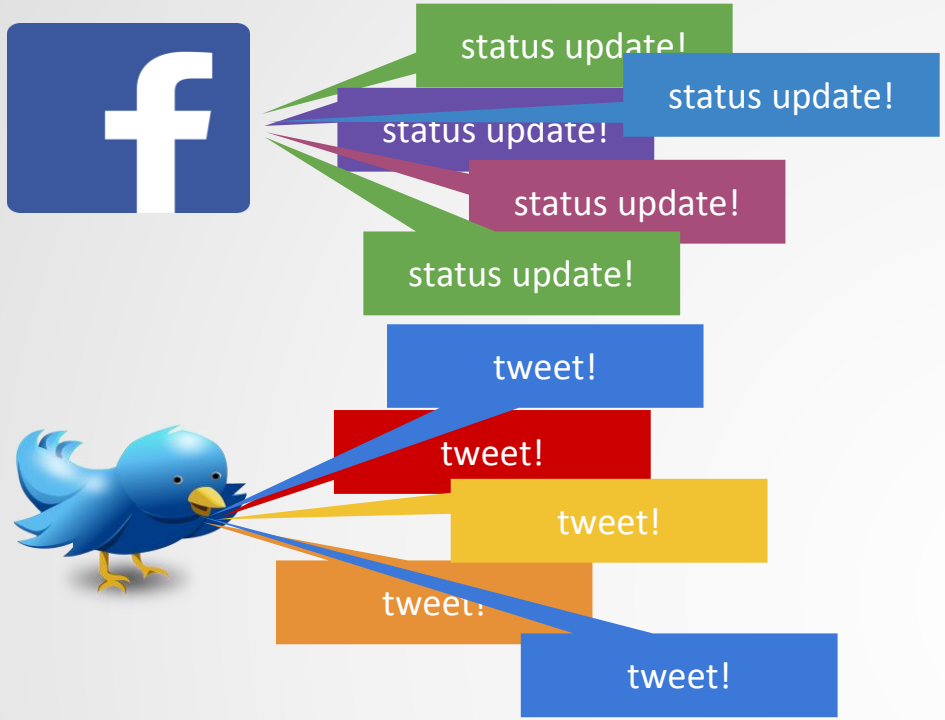
LIWC Dictionaries

Category	Example Words
Achievement	accomplish, beat, master, plan, quit
Articles	a, alot, an, the
Body	feet, goosebumps, skin, head, hands
Causation	makes, origin, rationale, used, why
Death	alive, bury, coffin, death, fatal, war
Filler	blah, like, oh well, you know, I mean

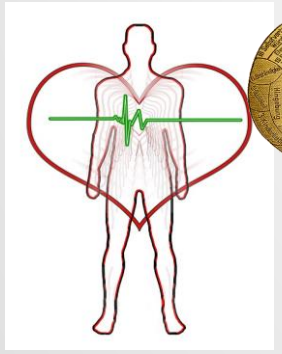


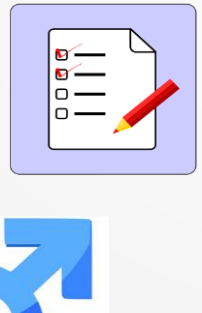

THE KEYS TO HAPPINESS

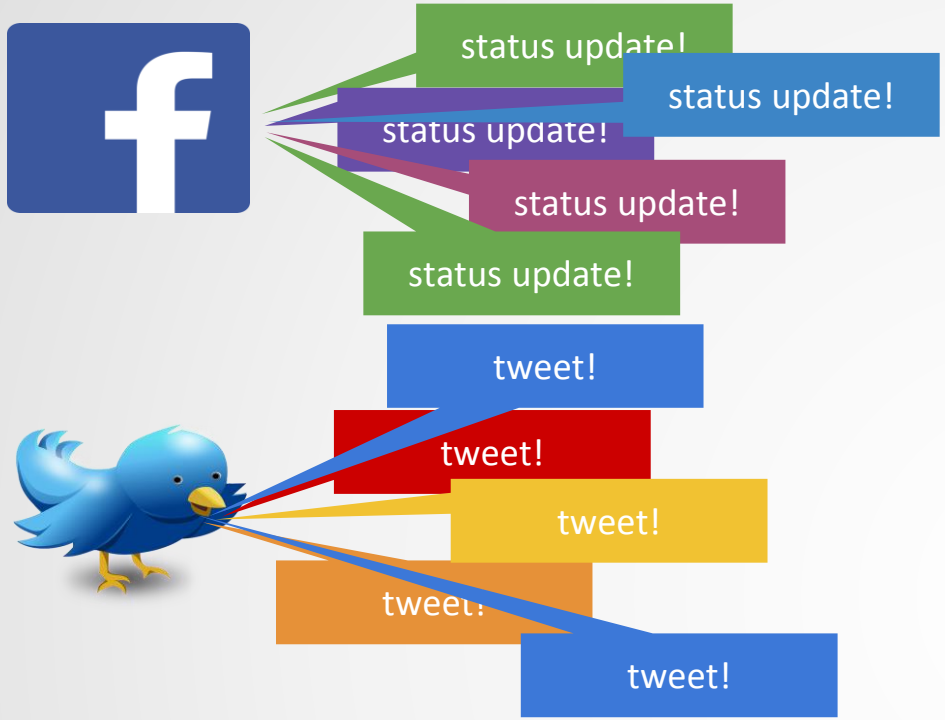

status update!
status update!
status update!
status update!
status update!
tweet!
tweet!
tweet!
tweet!
tweet!


status update!
status update!
status update!
status update!
status update!
tweet!
tweet!
tweet!
tweet!

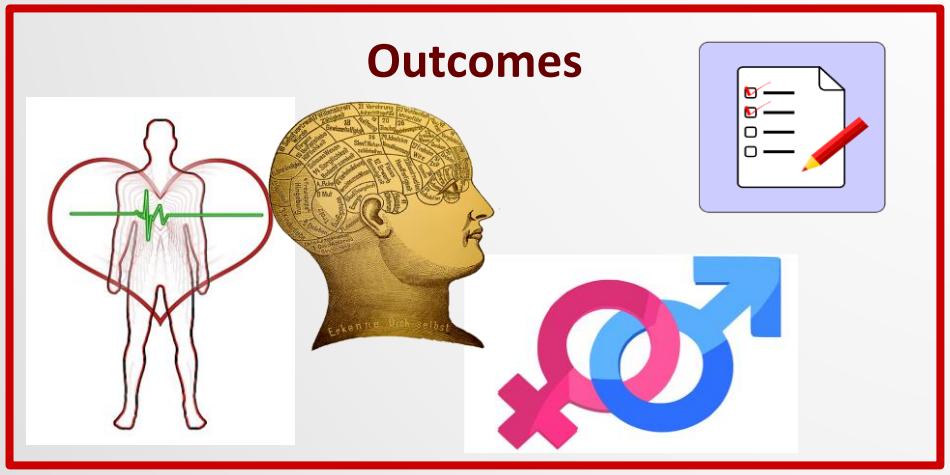
Outcomes





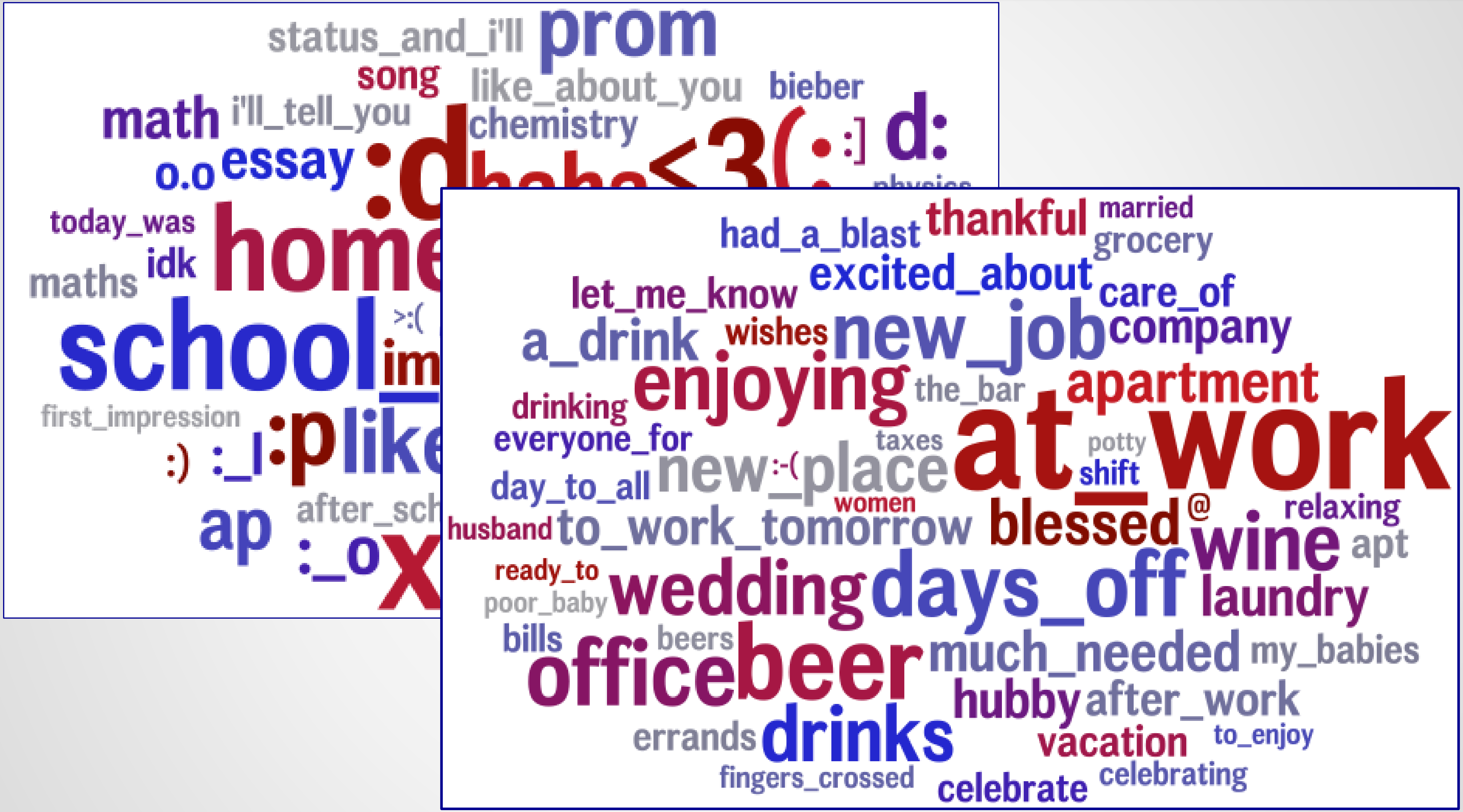


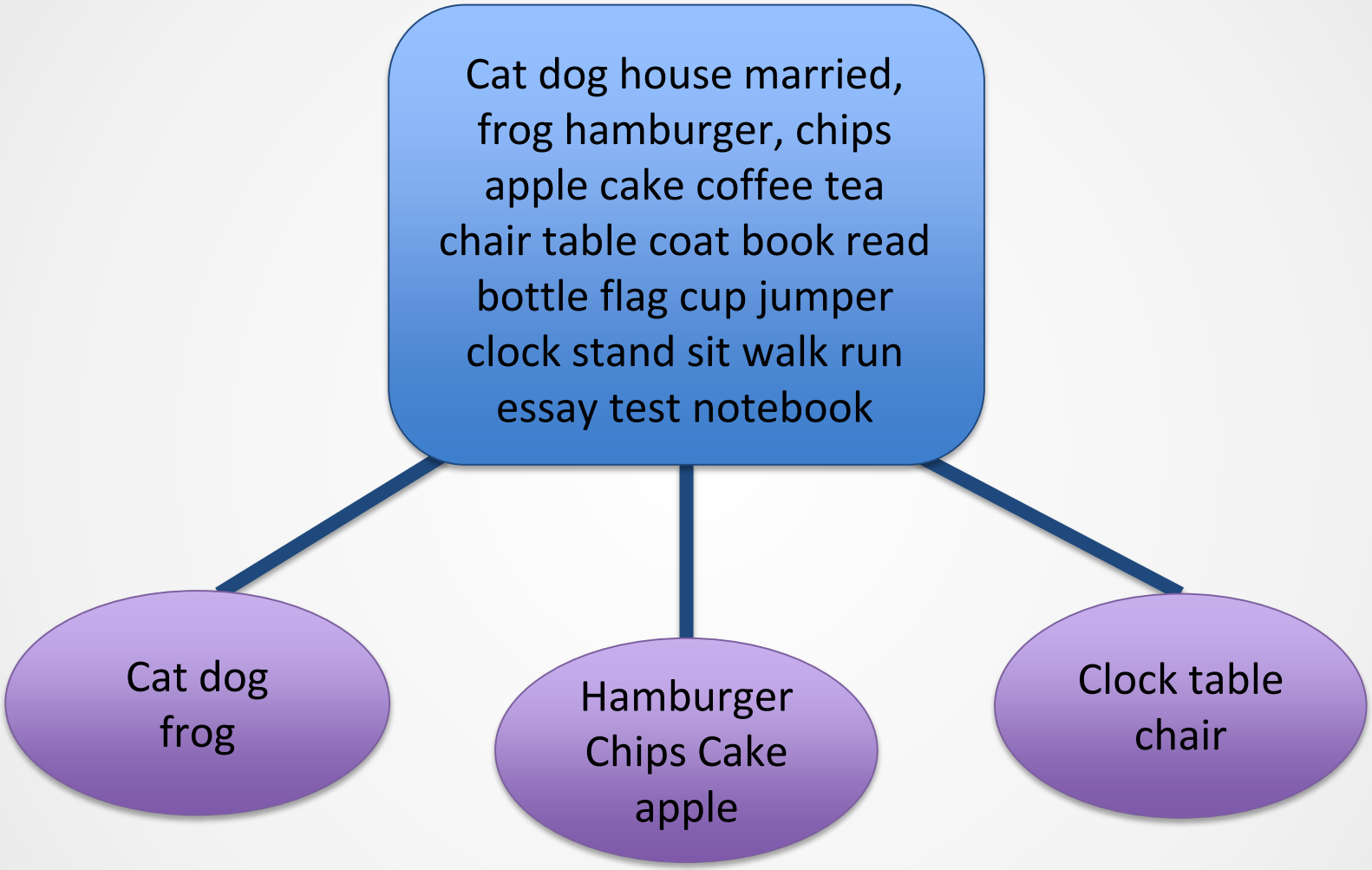
language analysis



Kern et al., 2014, *Assessment*

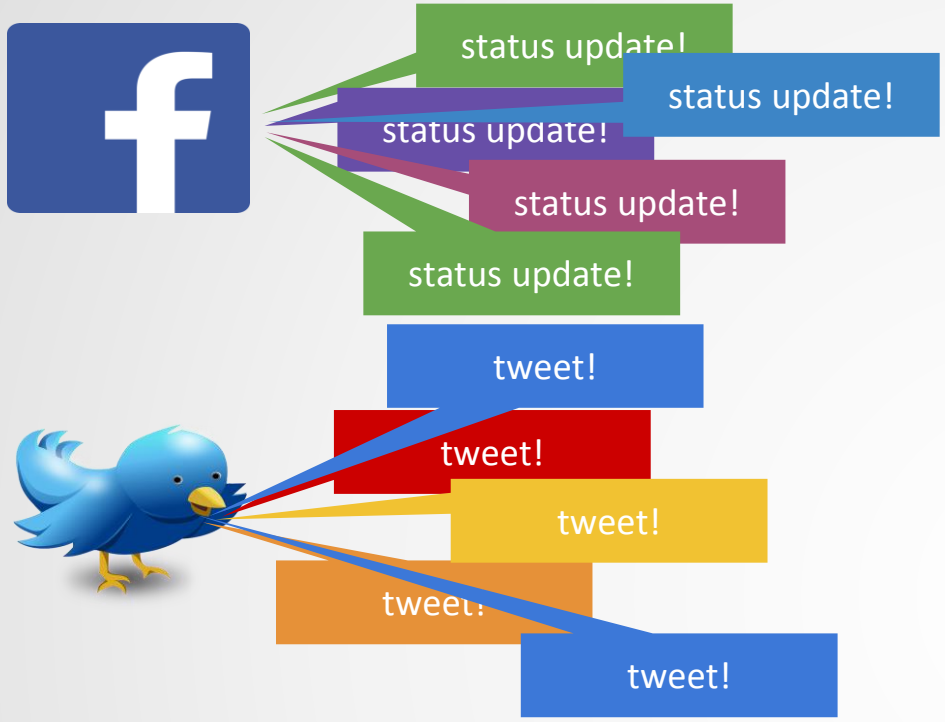
Kern et al., 2014, *Assessment*


status_and_i'll
prom
song
like_about_you
bieber
math
i'll_tell_you
chemistry
:d
haha
<3
(:
:]
d:
o.o
essay
today_was
home
idk
maths
school
>:(
first_impression
:p
like
:)
:_l
ap
:_o
had_a_blast
thankful
married
grocery
excited_about
let_me_know
care_of
a_drink
wishes
new_job
company
enjoying
the_bar
apartment
drinking
at_work
everyone_for
taxes
potty
new_place
:-(
shift
day_to_all
women
relaxing
@
husband
to_work_tomorrow
blessed
wine
apt
ready_to
wedding
days_off
laundry
poor_baby
bills
beers
office
beer
much_needed
my_babies
hubby
after_work
errands
drinks
vacation
to_enjoy
fingers_crossed
celebrate
celebrating


Cat dog house married, frog hamburger, chips apple cake coffee tea chair table coat book read bottle flag cup jumper clock stand sit walk run essay test notebook
Cat dog frog
Hamburger Chips Cake apple
Clock table chair

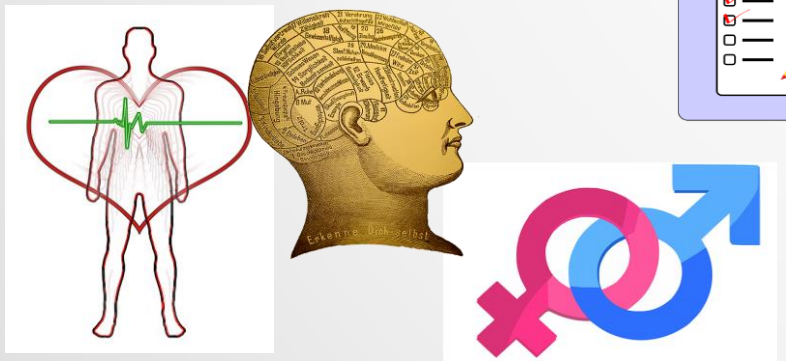

very_busy
fairly
enjoy
lead
great
very_interesting
love
very_much
life
busy
moment
interesting
hobbies
find
spend
b
b
b
prevalence in topic

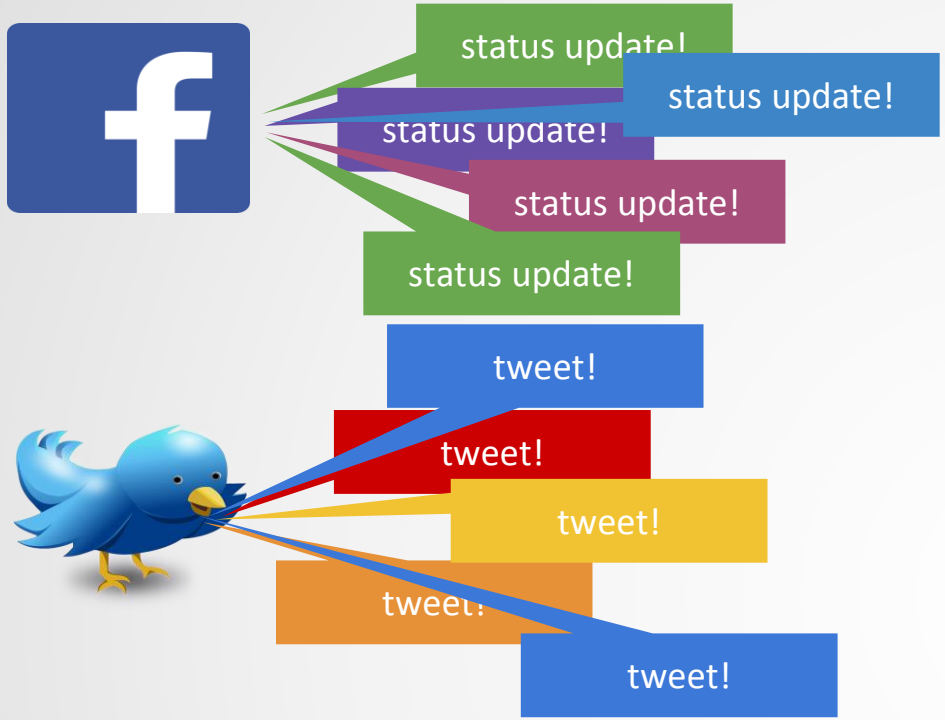
Participant 1	Participant 2	Participant 3
topic 1: .05	topic 1: .03	topic 1: .04
topic 2: .02	topic 2: .01	topic 2: .03
topic 3: .01	topic 3: .03	topic 3: .03
…	…	…
topic 100: .07	topic 100: .05	topic 100: .06



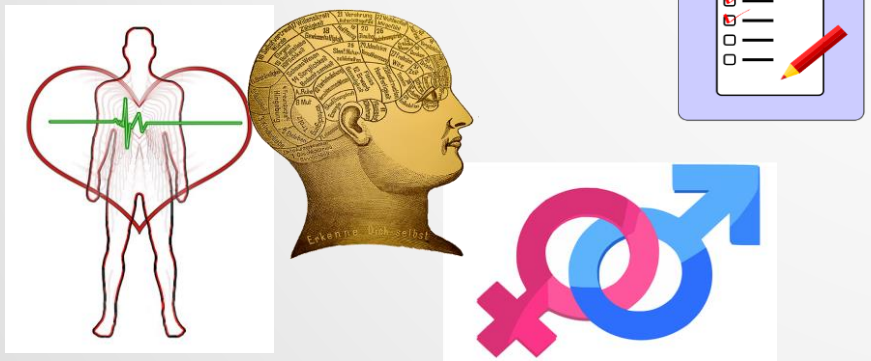
Words
Phrases
Topics

Outcomes





Outcomes



Words
Phrases
Topics

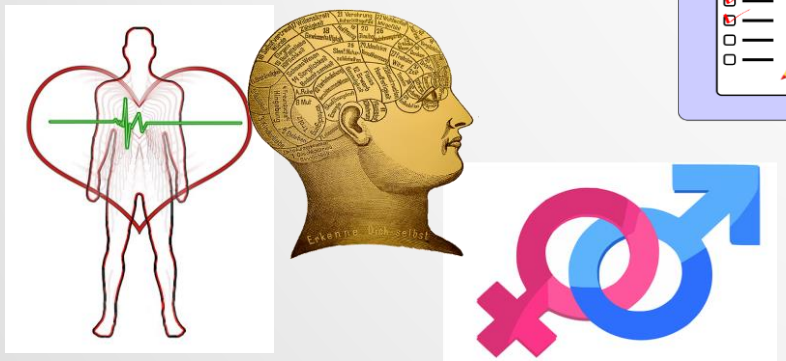
Prediction

Exploration


WEBSTER'S II

Words
Phrases
Topics

Outcomes



Prediction

Exploration

1958	1965	1969	1974	1981	1991
Birth	7	11	16	23	33
17,415	15,425	15,337	14,654	12,537	11,469
2000	2003	2004	2008	2013	
42	44/45	46	50	55	
11,419	9,377	9,534	9,790	9,137	

Imagine you are now 25 years old. Write about the life you are leading, your interests, your home life, and your work at the age of 25.

7. Imagine that you are **now** 25 years old. Write about the life you are leading, your interests, your home life and your work at the age of 25. (You have 30 minutes to do this).

I am 25 years of age I work in [redacted] that is a gorge in [redacted] road. I live in [redacted] road number 90, so do not have a lot [redacted] worke geting there. I go by my mini copper S. with all my home made words and lights & nobs. [redacted] when I get to work it is about 8 o clock, [redacted] I get out of bed about 5 min to 8 clock, get dress and have a glass of milk and go to work. after work I go to the cave, and have a cocol and a romreg. then I go down the

Do not write here.

Col. 51 C

Col. 52

Col. 53 4

Col. 54

Col. 55

Col. 56

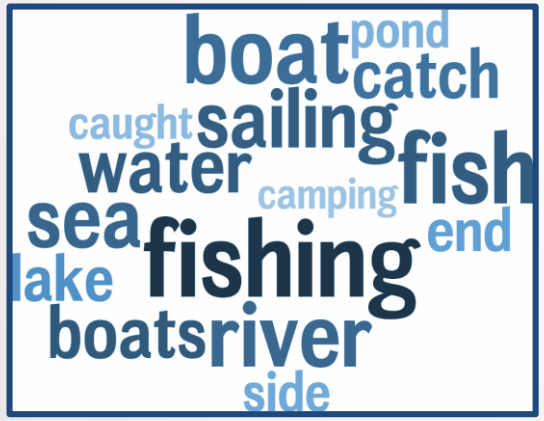
Do not write here.

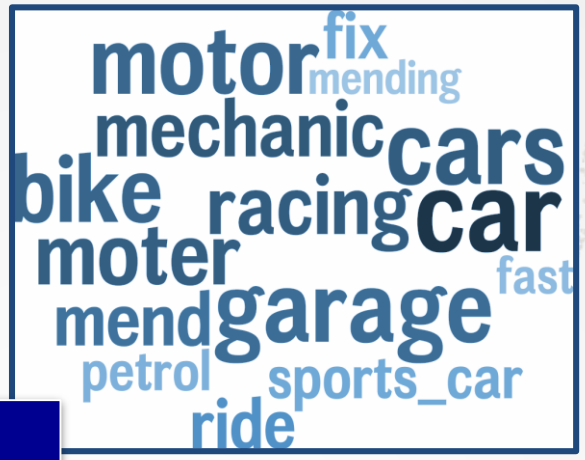
bowl to have a game. I take my own bowling ball and shoes, bec-ouse the shoes are to hard. and my fingers wont worght fit inx the ball. in the [redacted] bowl the have lones to bowl in. then it is about 7 o-clock, ~~and~~ I go~~t~~ home and have tea. Then I go to a dance in re~~a~~ding or go to london. then itis saterday, I work saturday morning, then go home for diner, and go to a football math in oxford. then watt tellevtoin or go out wiht the wife. to the rose inn or out some were exsiting. we have

Males


tennis
team
football_club
golf
watch
sports
cricket
football
play
matches
play_football
playing
club
sport
rugby

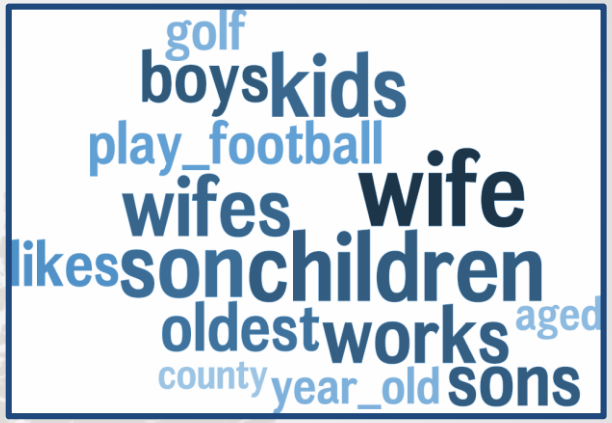

mini
van
test
driving
red
garage
morris
am_twenty_five
number
cortina
drive
ford
blue
car
capri

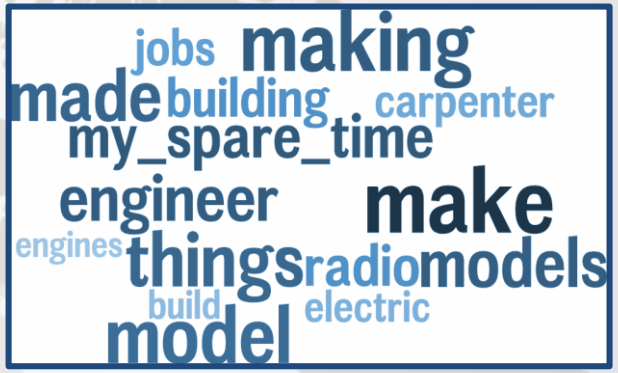

boat
pond
catch
caught
sailing
water
fish
camping
sea
end
fishing
lake
boats
river
side


motor
fix
mending
mechanic
cars
bike
racing
car
moter
fast
mend
garage
petrol
sports_car
ride


fast
rolls
cars
run
driver
drive
won
lorry
races
long
track
driving
race
racing
win

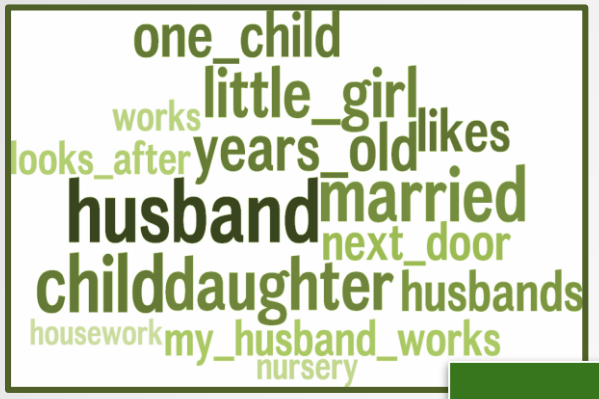

guns
gun
leave
tank
joined
training
camp
men
army
join
tanks
shooting
learn
fight
war

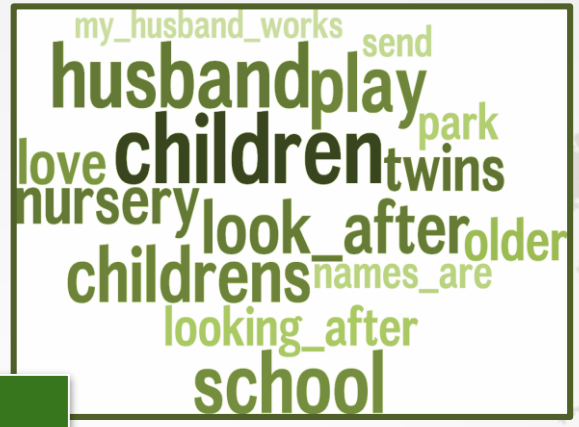

golf
boys
kids
play_football
wife
wifes
likes
son
children
oldest
works
aged
county
year_old
sons


jobs
making
made
building
carpenter
my_spare_time
engineer
make
engines
things
radio
models
build
electric
model


force
raf
test
aircraft
plane
airport
flying
planes
fly
pilot
air
aeroplane
training
flight
jet

Females


one_child
little_girl
works
likes
looks_after
years_old
husband
married
next_door
childdaughter
husbands
housework
my_husband_works
nursery


my_husband_works
send
husband
play
park
love
children
twins
nursery
look_after
older
childrens
names_are
looking_after
school


cook
sewing
interests
clothes
my_spare_time
watching
very_much
making
swimming
dancing
love
knitting
cooking
enjoy
reading

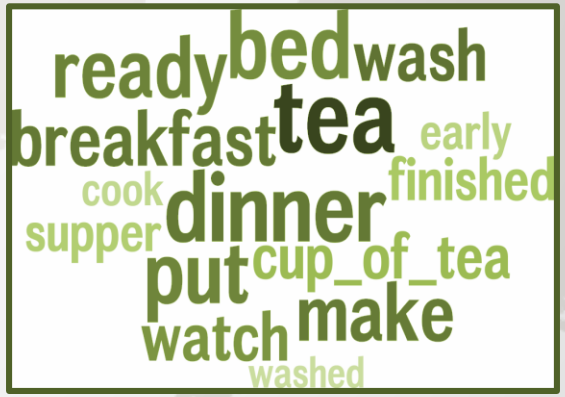

class
school
history
boys
teaching
primary
teacher
junior
learn
maths
teachers
english
girls
teach
games


give
husband
walk
sit
sleep
girl
look_after
park
baby
child
pram
babies
sitter
babys
boy

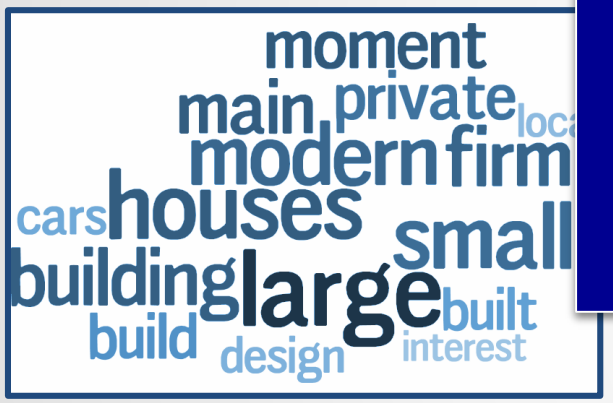

hairdressers
hairdressing
dressers
girls
dressing
set
people
cut
hairdresser
peopleshair_dresser
salon
lady
shop
hair

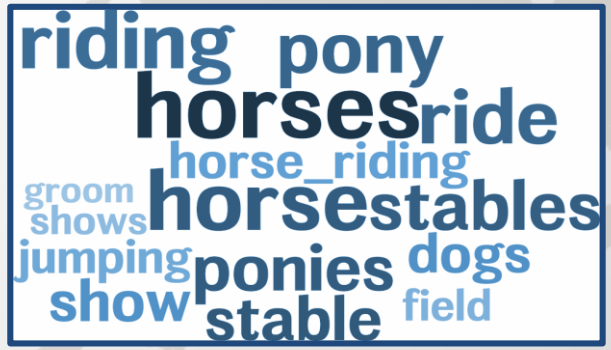

boyboyscalled
twinstwo_boys
children
call
names
boy's
two_girls
twin
older
girl
girls
marry


sister
nurse
doctors
nursing
ward
nurses
hospital
doctor
docter
childrens
training
sick
ill
duty
patients

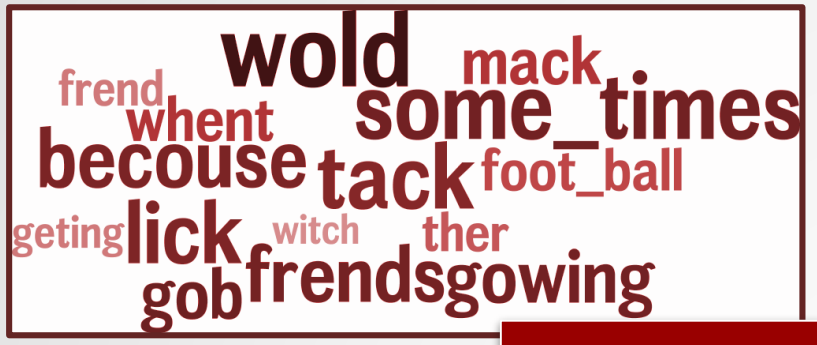

ready
bed
wash
breakfast
tea
early
cook
finished
supper
dinner
put
cup_of_tea
watch
make
washed

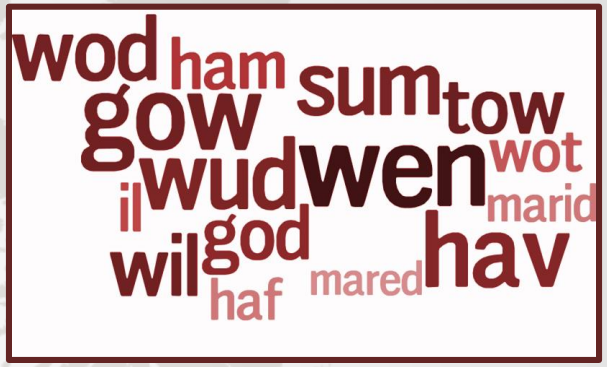
Higher social class

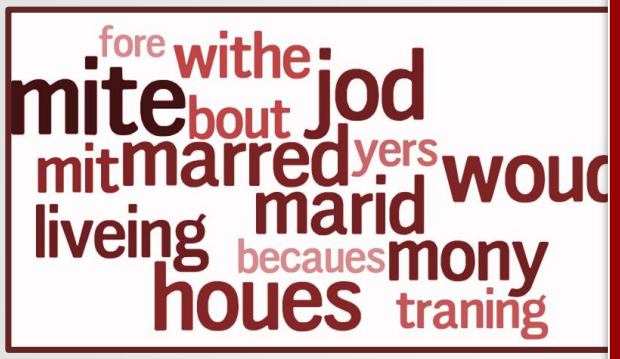


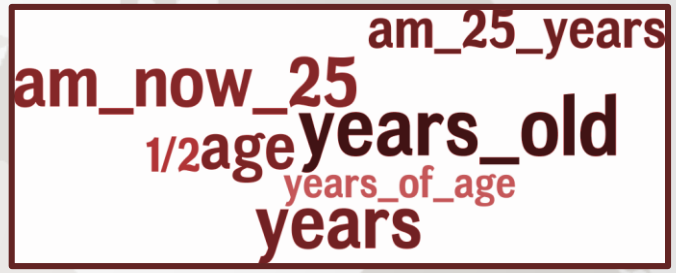


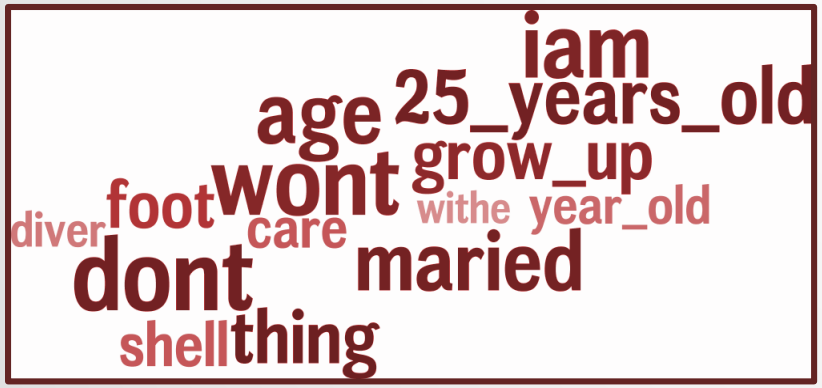
Lower social class











Research Questions

- Do essay linguistic features predict future outcomes across five decades, in a nationally representative sample?
- What part of the variance overlaps with childhood survey data?
- What can we discover that is unexpected from the essays?

Overview

- **Health**: What can we learn from the language used by 11 year olds about their future health and well-being? (Margaret Kern)
- **Physical activity**: Language as a predictor of lifetime physical activity (Benedetta Pongiglione)
- **Cognition**: Language as predictor of cognitive functiontioning (Martina Narayanan)
- **Social mobility:** Social mobility, and ‘which dreams came true’? (Alissa Goodman)

Using Childhood Essays to Predict Physical and Mental Health Across the Lifespan

Margaret L Kern, H. Andrew Schwartz,
Alissa Goodman, Martina Narayanan

CENTRE FOR LONGITUDINAL STUDIES

THE UNIVERSITY OF MELBOURNE

Stony Brook University

www.peggykern.org

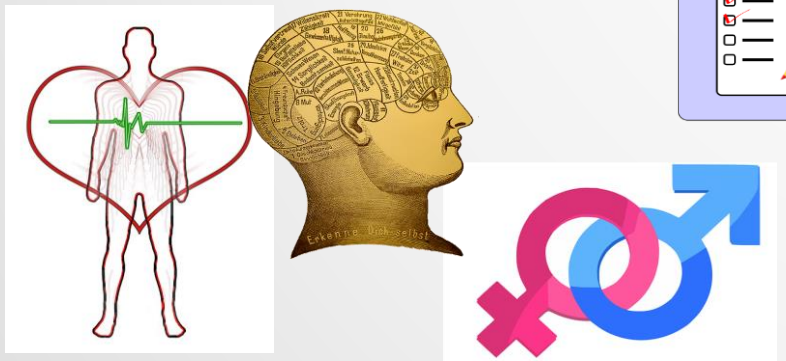
Peggy.Kern@unimelb.edu.au

@pkern001


9
8
7
6
5
4
3
2
1

Words
Phrases
Topics

Health Outcomes



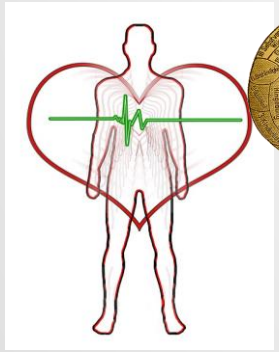
Prediction

Exploration

Words
Phrases
Topics

Prediction
Exploration

Health Outcomes




Getty

17,415 baseline cohort

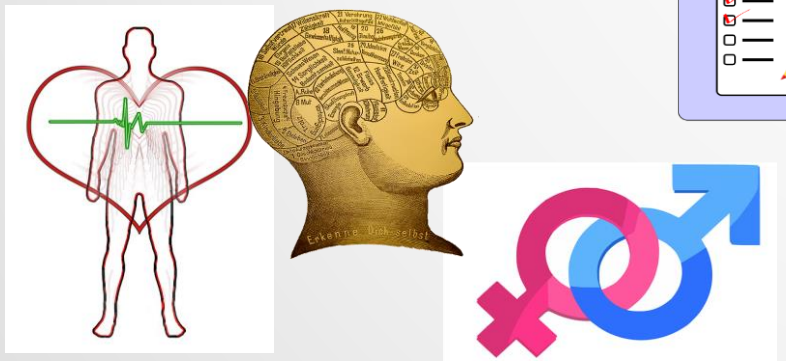
15,337 year 11 surveys

10,500 essays transcribed

Age 23	Age 33	Age 42	Age 44	Age 50	Age 55
7,822	7,110	7,140	5,790	6,130	5,778

Words
Phrases
Topics

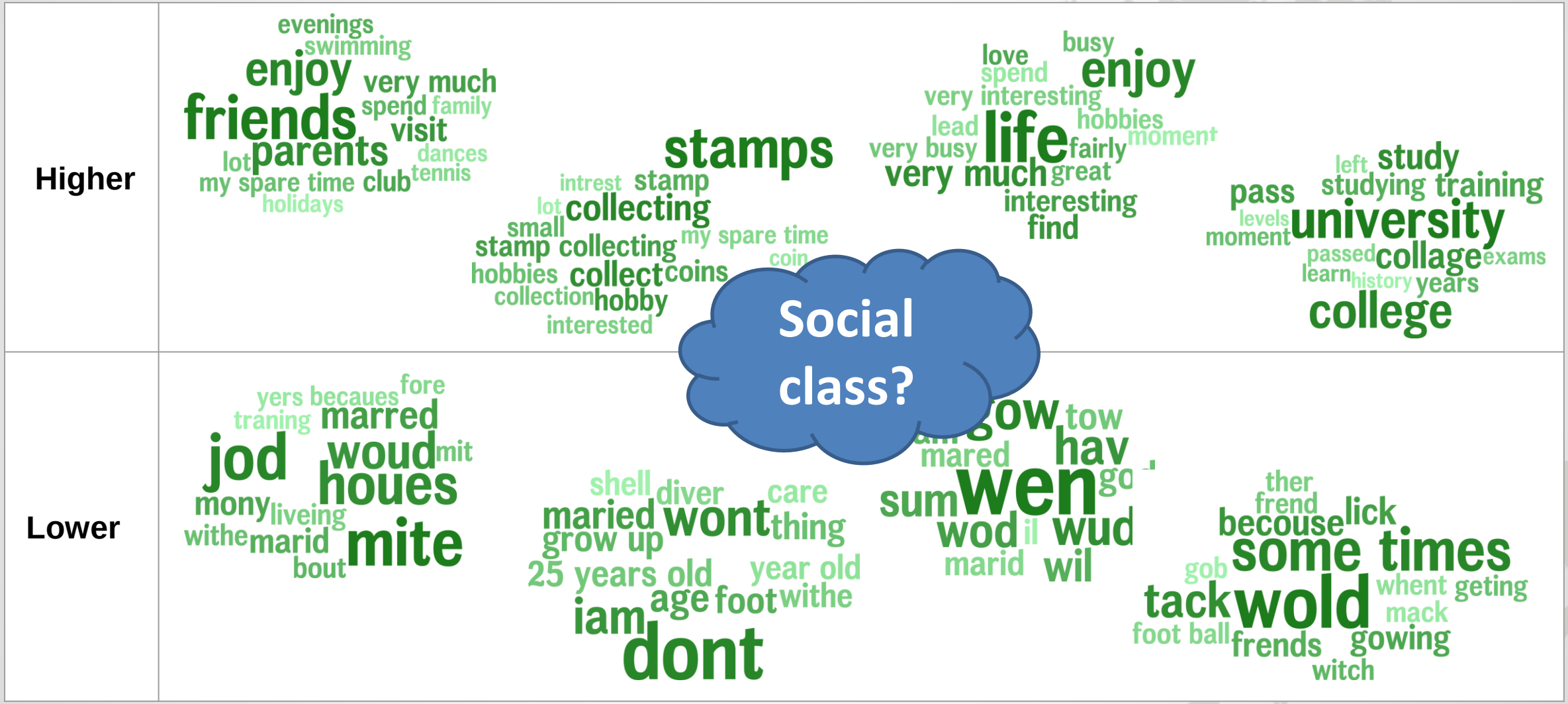
Health Outcomes



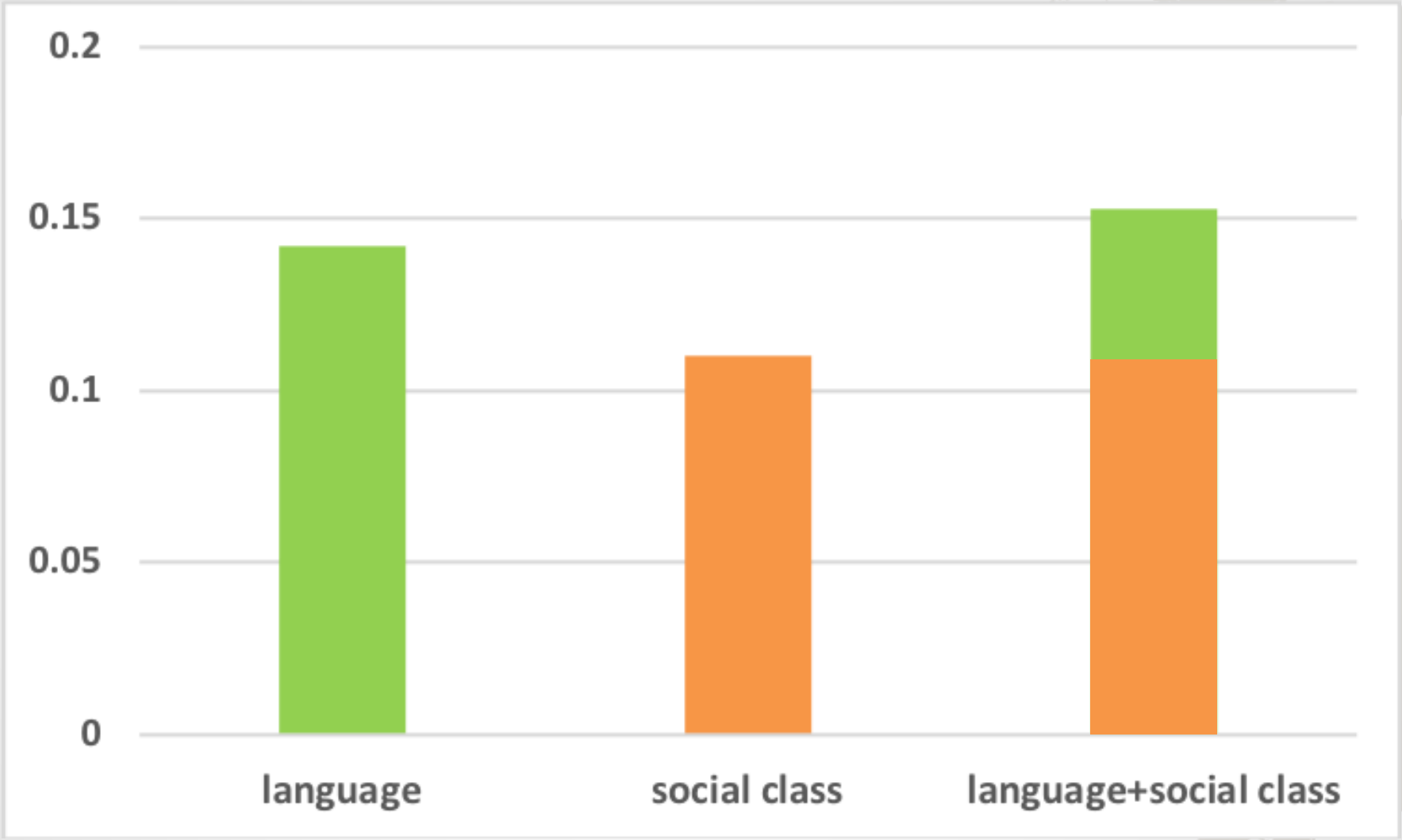
Prediction

Exploration

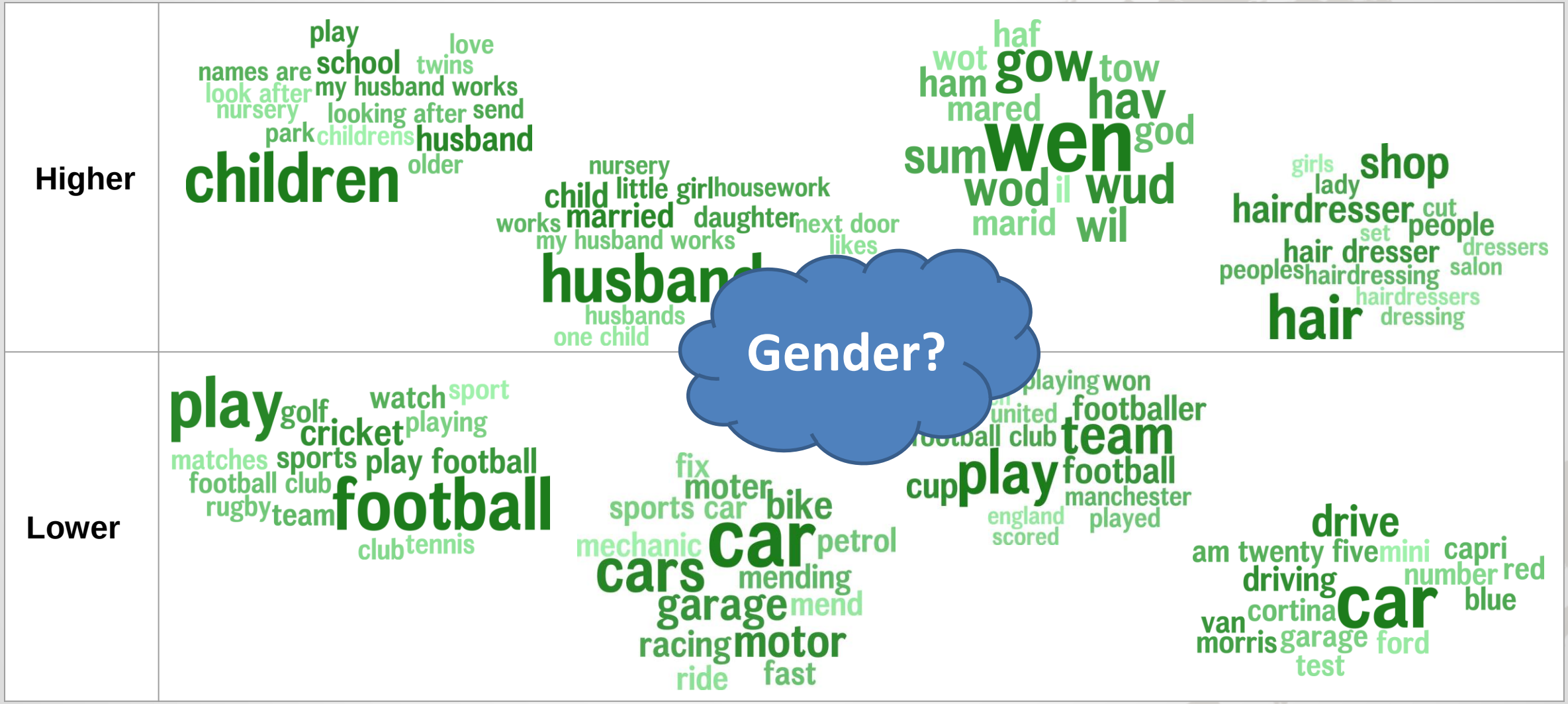
Topics consistently predicting self-rated health



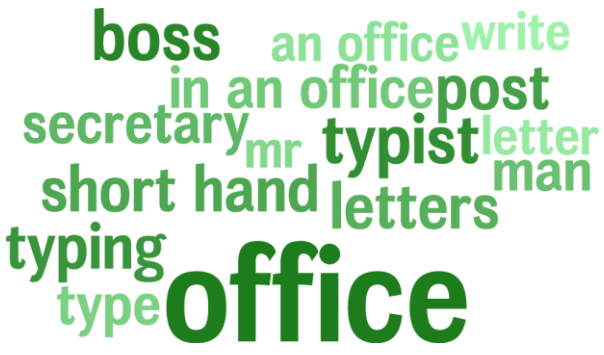
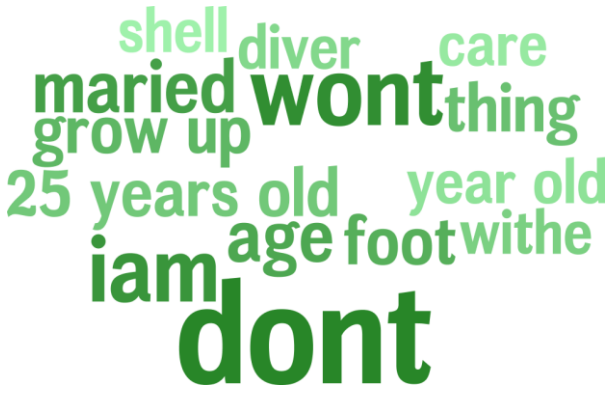
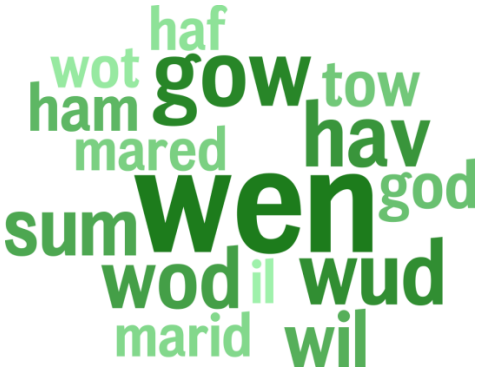
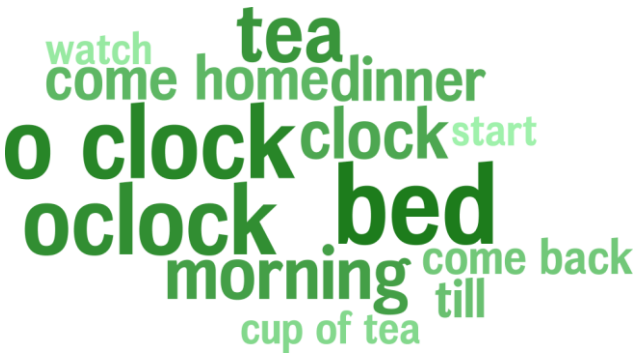
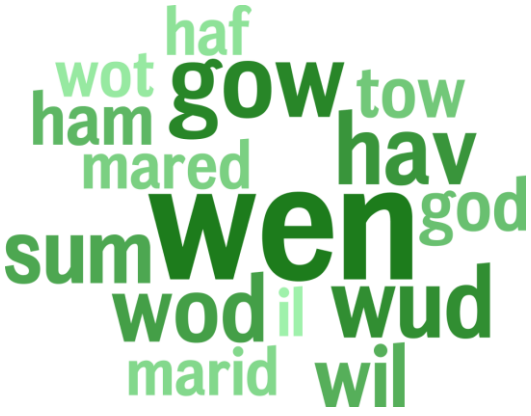
Self-Rated Health Predictive Accuracy



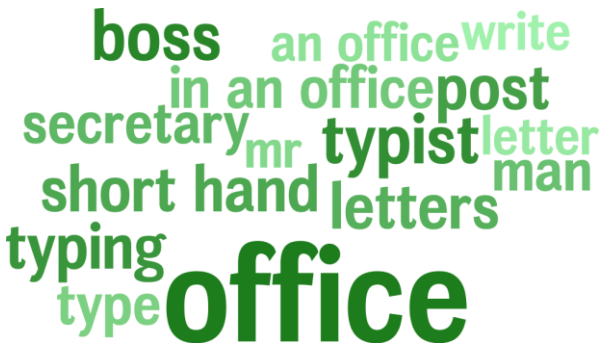
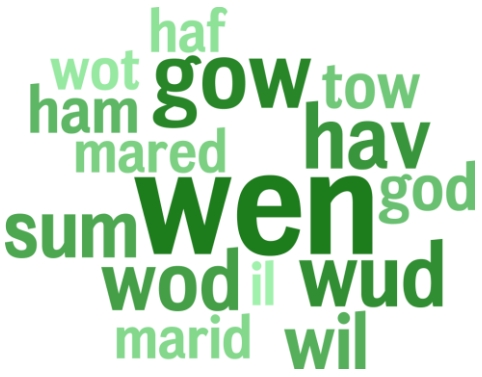
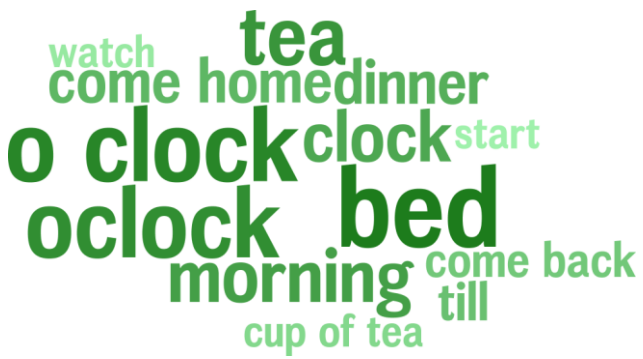
Topics consistently predicting malaise



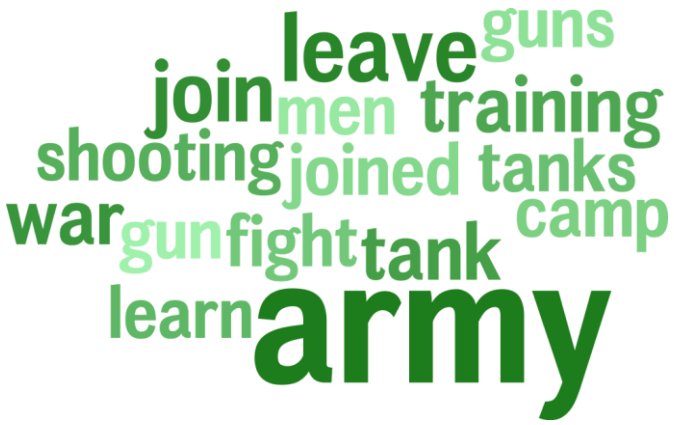
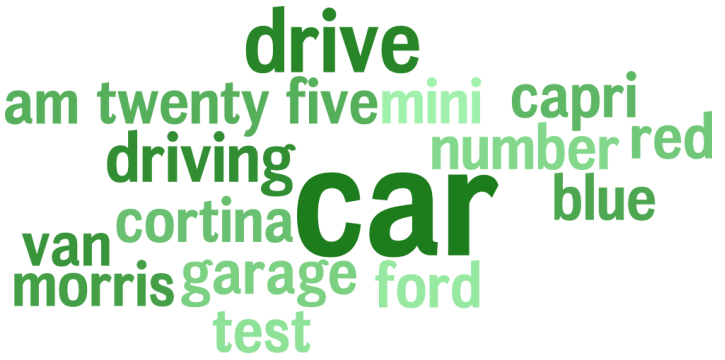
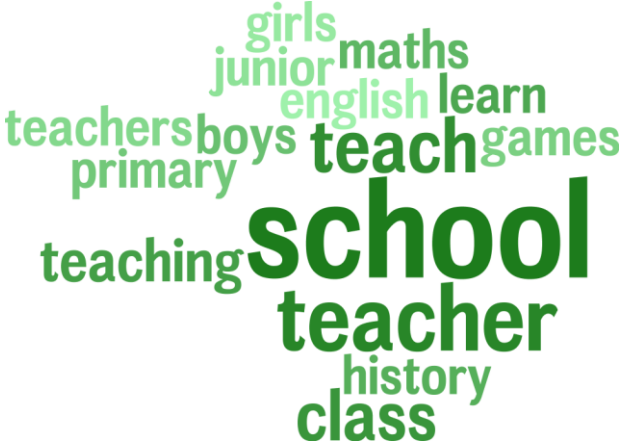
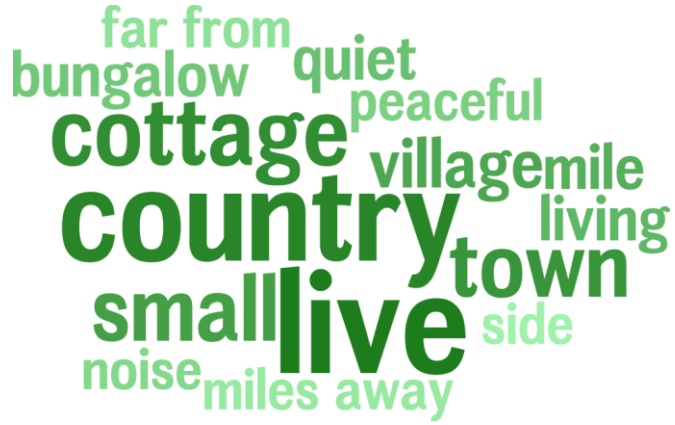
Topics predictive of age 50 malaise by gender

males			
females			

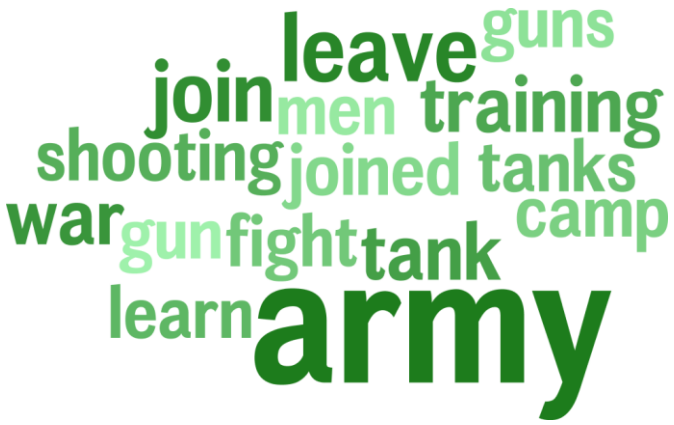
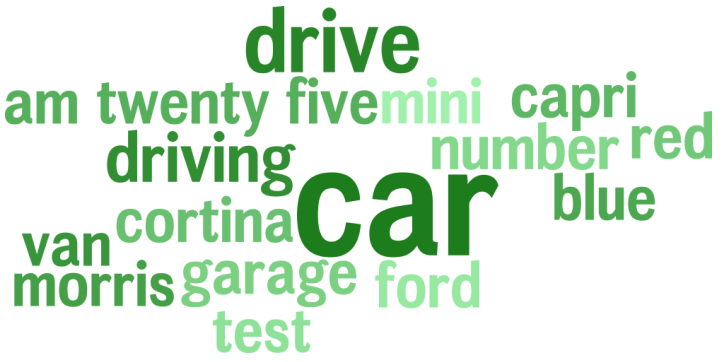
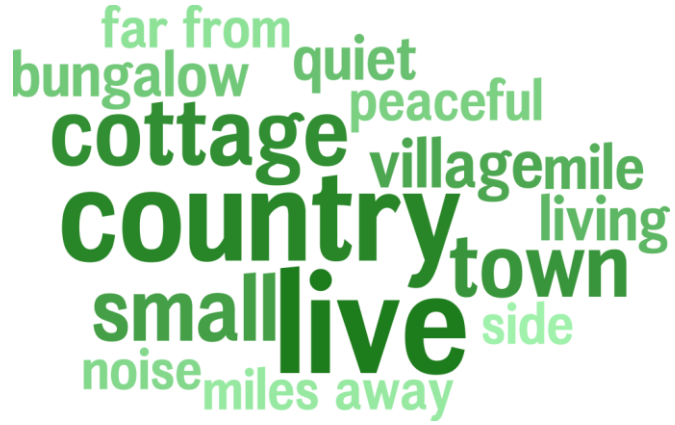
Topics predictive of age 50 malaise by gender

males			
females			

Topics predictive of age 50 non-malaise by gender

males			
females			

Topics predictive of age 50 non-malaise by gender

males			
females			

Thank you!!

Margaret L Kern, H. Andrew Schwartz,
Alissa Goodman, Martina Narayanan

www.peggykern.org Peggy.Kern@unimelb.edu.au @pkern001

Possible and actual selves: Do children's expectations about future physical activity predict their adult activities?

JD Carpentieri, B. Pongiglione, M. L. Kern, A. Goodman,
N. Gupta & H.A. Schwartz

60 Years of our Lives Conference

London, 9 March 2018

Outline

- Background & Aims
- Methods
- Results
- Discussion

Outline

- Background & Aims
- Methods
- Results
- Discussion

Background

- Regular physical activity (PA) contributes to a range of positive health outcomes
- But a large proportion of the population does not meet PA recommendations

PA tracking over the life course

- Long-term studies show low levels of PA tracking over the life course
  - National Child Development Study (NCDS): Significant but modest association between PA at 16 and 42
  - British Cohort Study 1970 (BCS70): Association between age 16 and 42 was weaker than in NCDS

PA tracking over the life course

- Tracking better for males than females
  - Females: lower overall participation rates and/or the impact of major transitions in the life course
- Understanding why PA levels change across the life course could potentially provide useful insights for the targeting of PA interventions

Possible Selves & Physical Activity Identity (PAI)

- People draw on their past and current selves to imagine a range of possible future selves
- Physical Activity Identity(PAI): importance of PA to one's conception of oneself
- Writing about leisure interests such as PAs may give children an opportunity to express continuity of identity over time (Elliott and Morrow, 2007)

Aim and Research Questions

- Investigate the predictive capacity of age-11 PA projections on lifetime physical activity
- Does mentioning PA in age-11 essays predict physical activity from age 33 onwards?
  - Does it predict PA at each time point?
  - Does it predict trajectory of PA?

Outline

- Background & Aims
- Methods
- Results
- Discussion

Methods

Data

- NCDS age 11, 23, 33, 42, 50 and 55
- Sweep 7 (age 46) excluded due to inconsistency with other sweeps

Sample

- NCDS respondents who completed age-11 essay (N= 10,567)

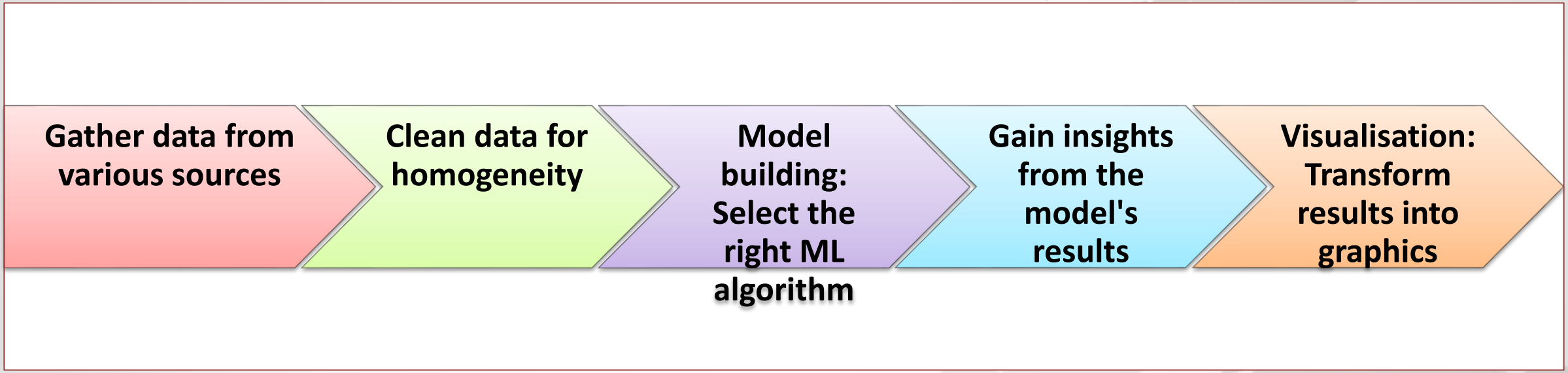
Measures

- **Outcome measures**
- Physical activity at different time points
  - Age 23: how often played sport in past 4 weeks
  - Age 33 to 55: frequency of regular exercising (1=once a week or more; 0=less than once a week)
- Cumulative score of PA across adulthood, ranging from 0 (never active) to 4 (always active) using ages 33, 42, 50, & 55
- Trajectories of PA: 4 class-model using latent class analysis

Exposure Measures

- Linguistic feature of PA: active and passive (i.e. as spectator) interest in physical activity'

Analysis: Machine Learning Process



Analysis: Physical Activity Predictions

Derived dictionary & frequencies

been	1692
wife	1674
football	1649
people	1643
were	1633
boy	1628
little	1618
place	1554
hope	1547
tea	1508
shop	1504
look	1496
off	1472

Activity Label	Model Accuracy	F-Score
Active	74.5%	0.74
Passive	87.2%	0.73

Analysis

- Association between linguistic feature of PA and outcomes:
  - At each sweep: logistic regression
  - Cumulative score of PA: linear regression
- Missing data: addressed using Multiple Imputation with chained equations

Outline

- Background & Aims
- Methods
- Results
- Discussion

The Spectator and the Active

- '[….] When I come home I just slump down in a chair and read a paper. When I have the chance I go to a football match [….]' (Spectator: Male, N22433X)
- '[….] my Job is a hair dresser. [….] I work from Eight o'clock in the morning to half past Eight at night. [….] When I go home at night I go swimming. for half an hour.[….]' (Active: Female, N24256G)

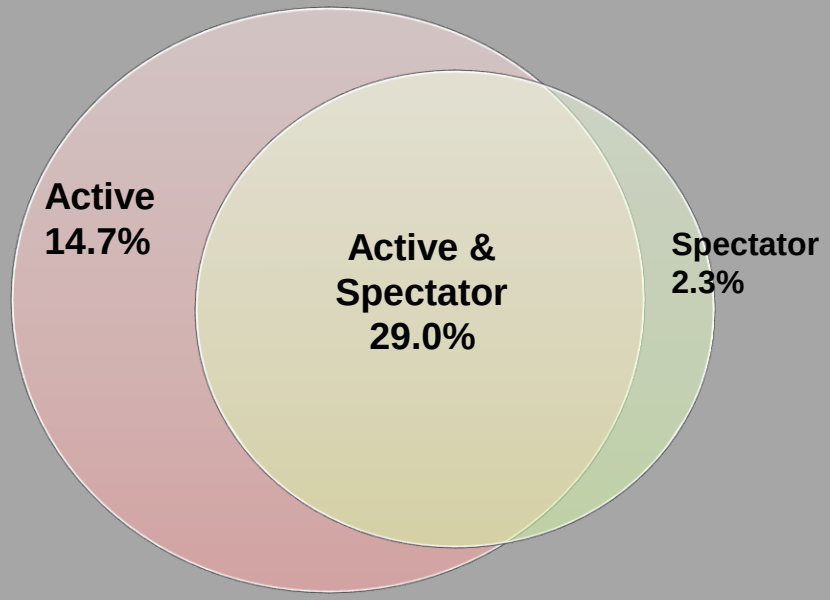
PA as part of projected daily routine

Intergenerational Continuity of PA

- ‘[….] when it is a hour to go to work I go out to play a game with my son Royston he is fond of football. He is in goal with the wind blowing toward me and I am in goal when it is blowing toward him [….]’ (Male, N24781Y)

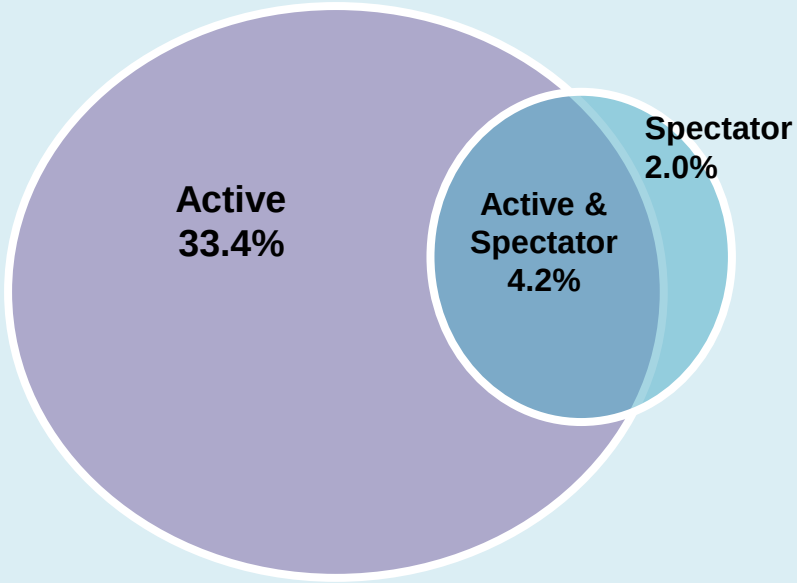
Linguistic Feature of PA

Males



Neither active not passive
54.0%

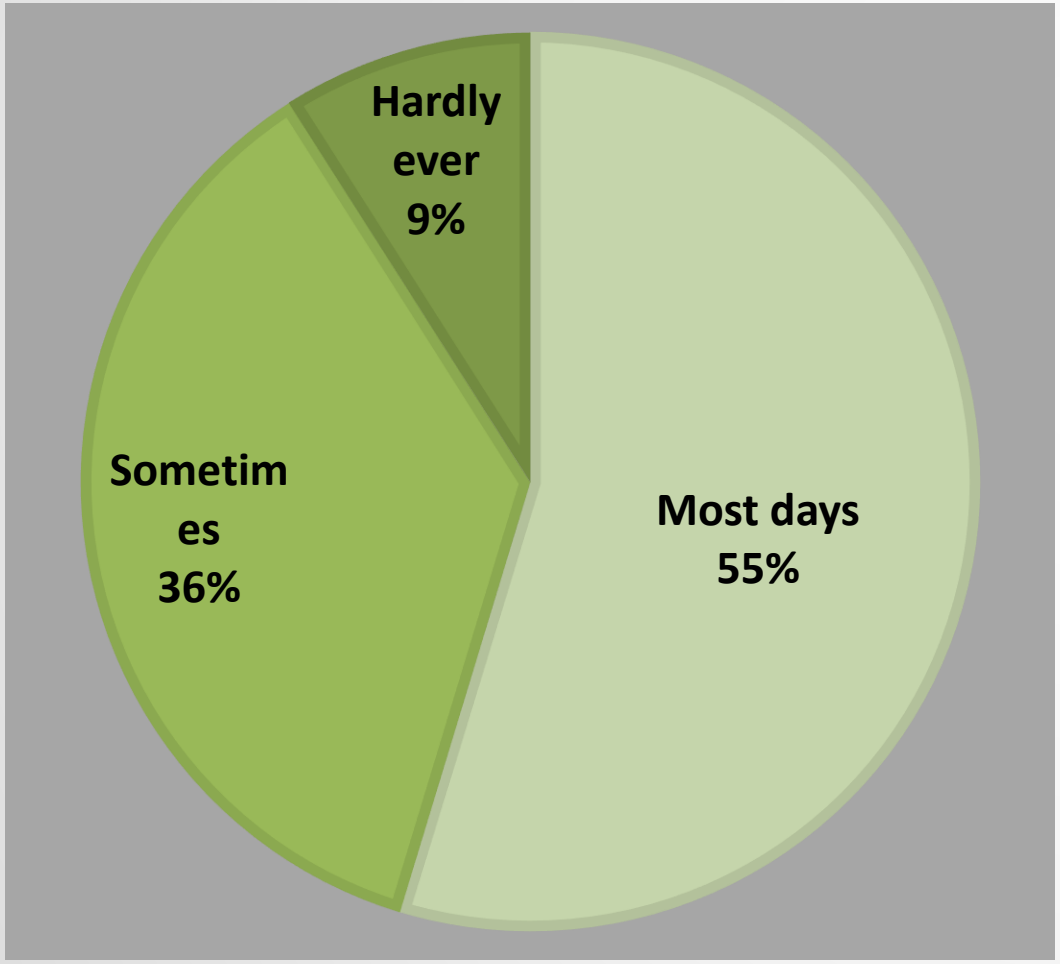
Females



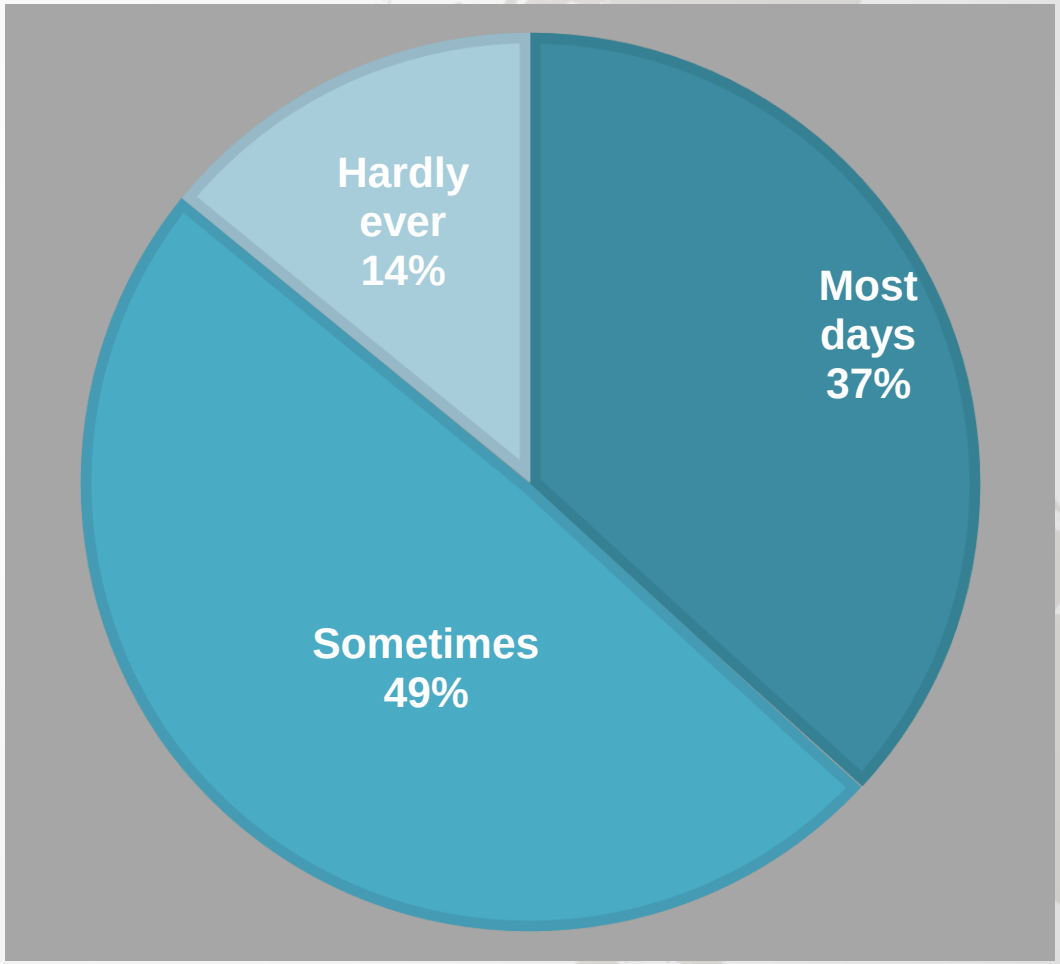
Neither active not passive
60.4%

PA measured in the survey

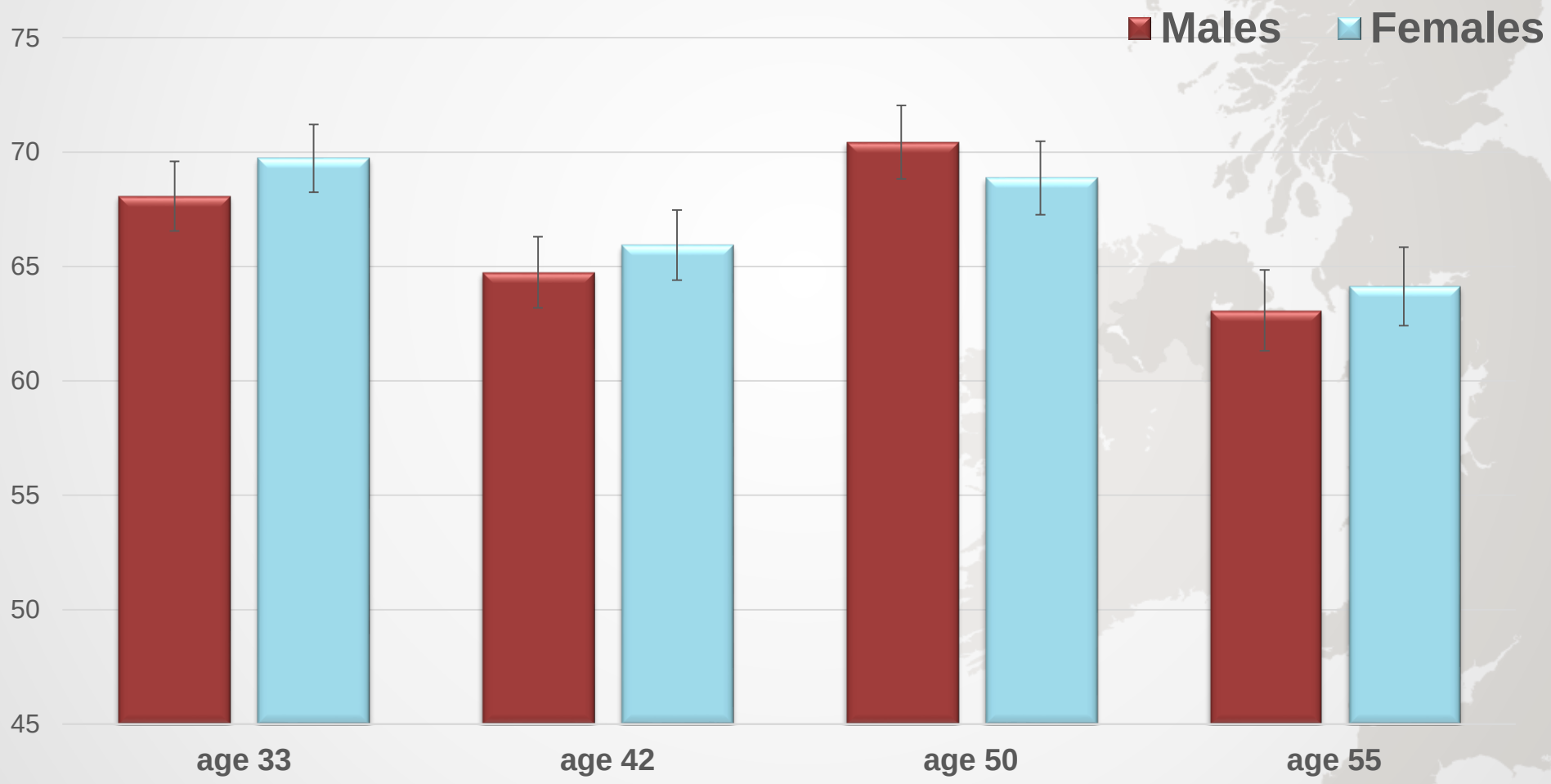
Males



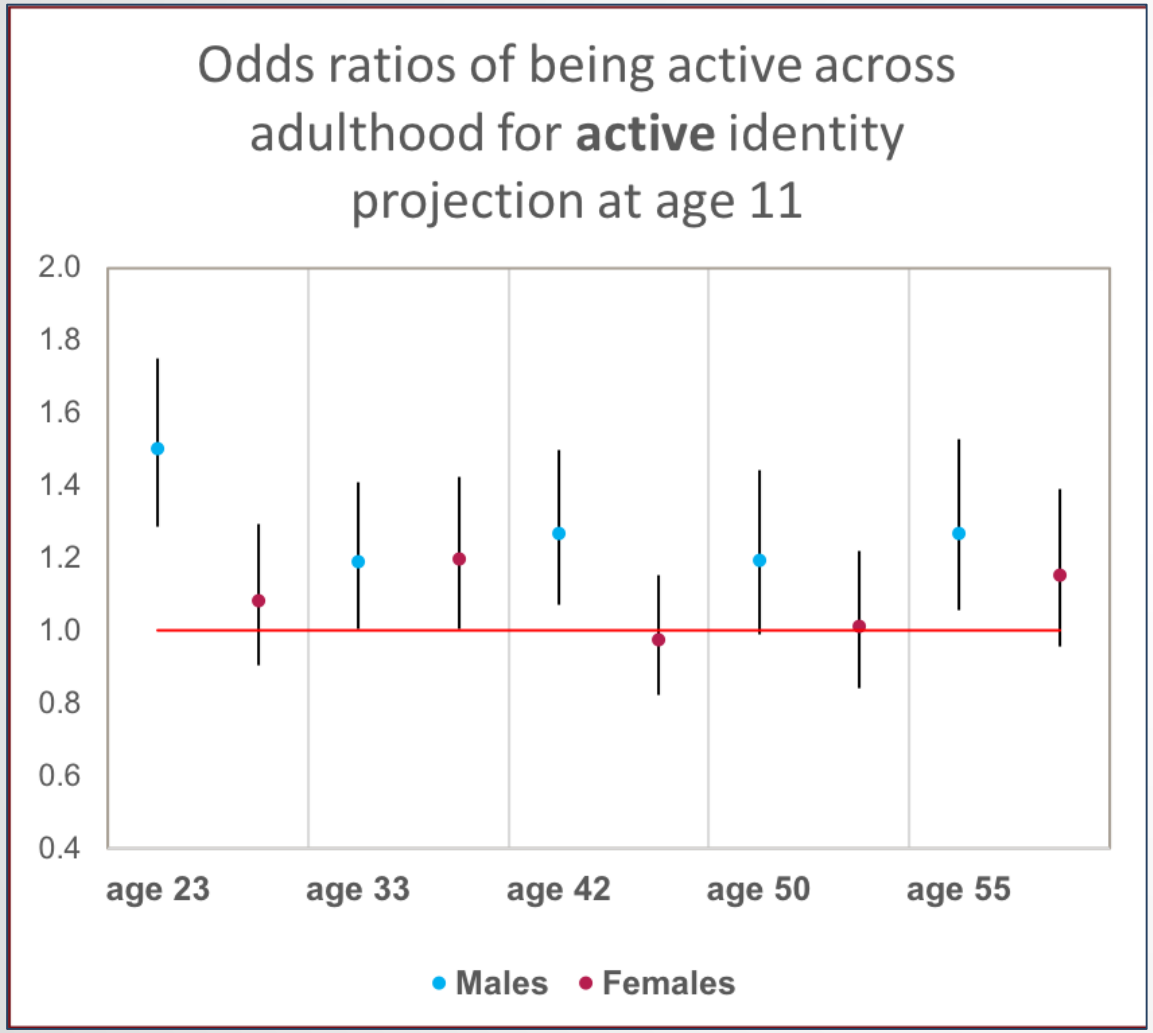
Females

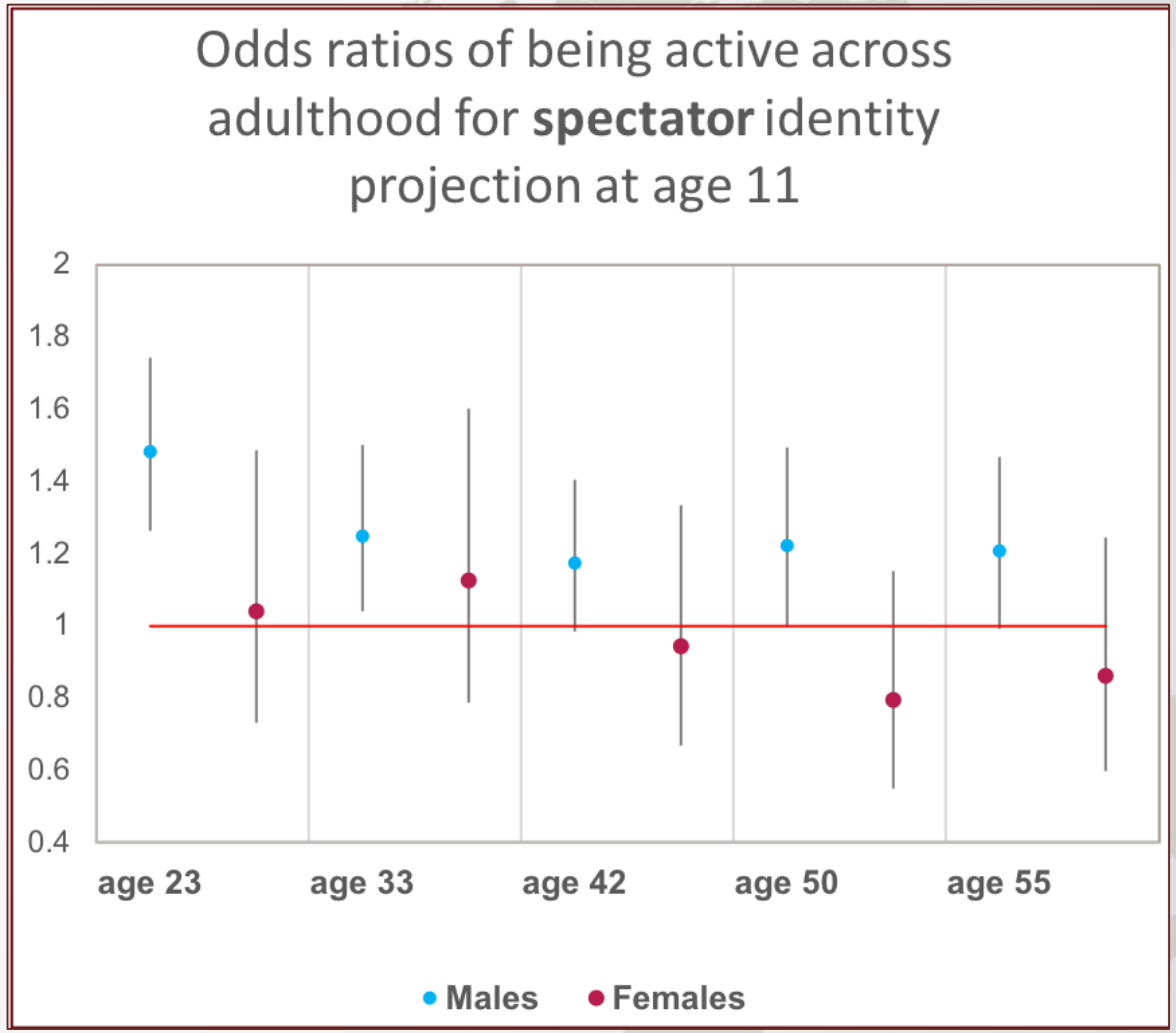


Proportion of respondents exercising once a week or more, across adulthood by gender

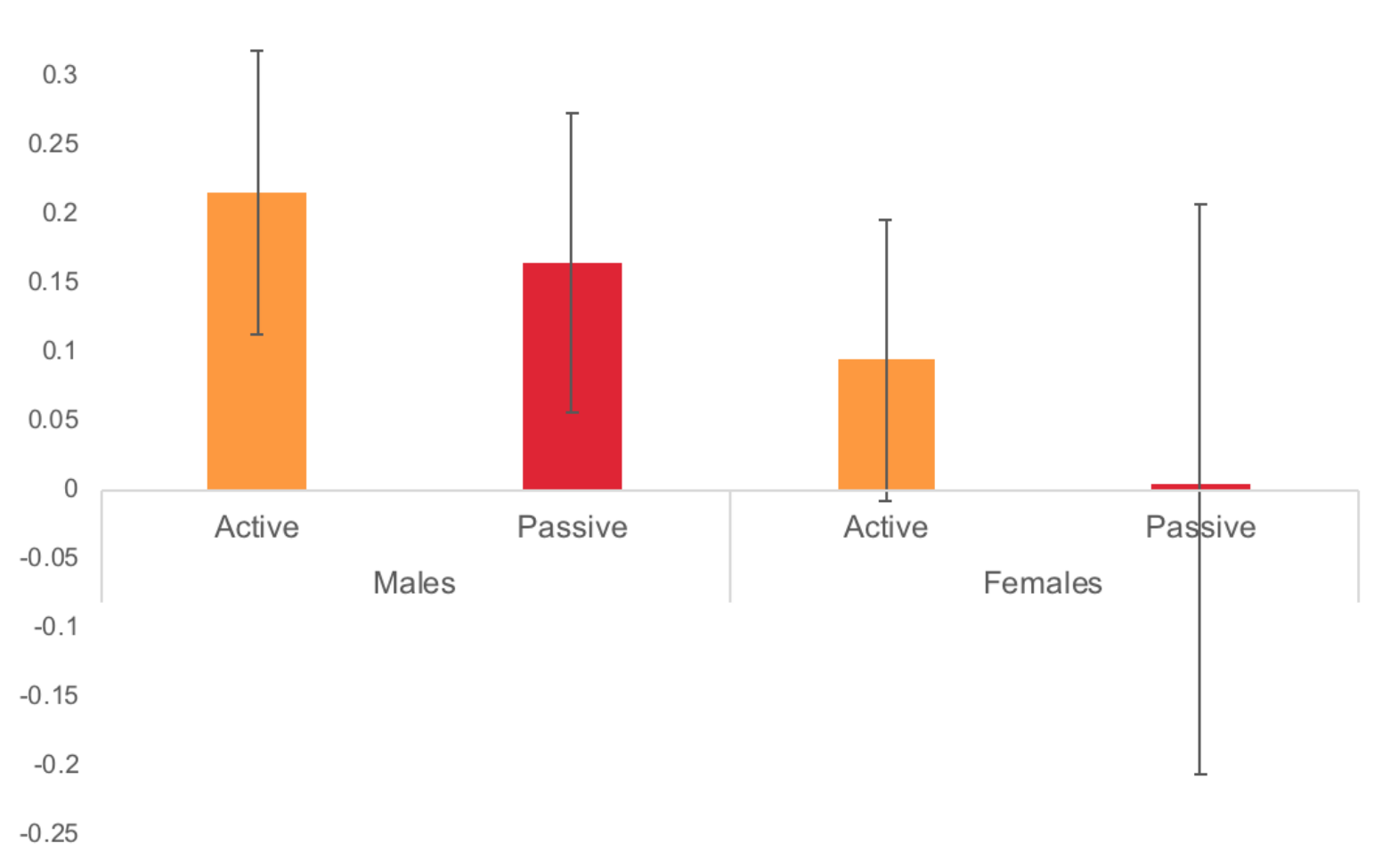


Results

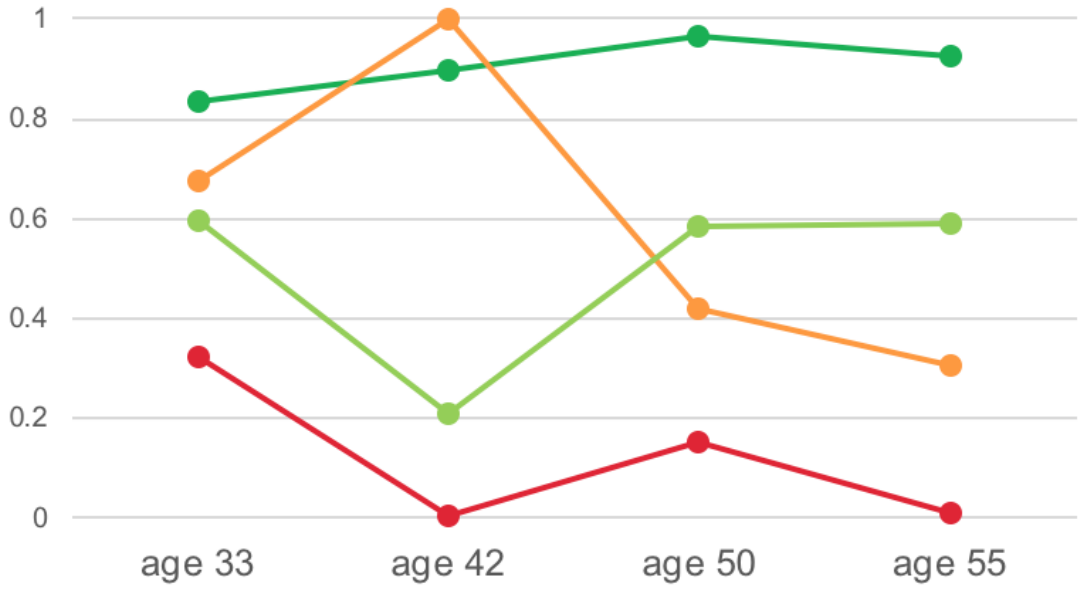

Odds ratios of being active across adulthood for **active** identity projection at age 11
2.0
1.8
1.6
1.4
1.2
1.0
0.8
0.6
0.4
age 23
age 33
age 42
age 50
age 55
Males
Females


Odds ratios of being active across adulthood for **spectator** identity projection at age 11
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
age 23
age 33
age 42
age 50
age 55
Males
Females

Association between linguistic feature of active and spectator identities of PA and cumulative score of PA, by gender

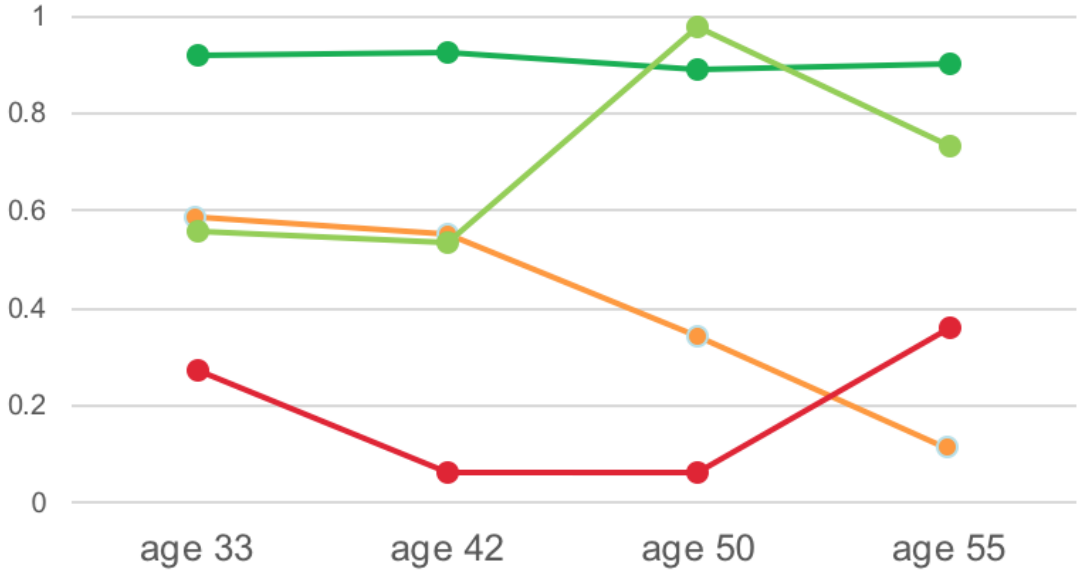


Trajectories of PA in adulthood, 4-class model, **women**



Always active	**45.6%**
Fluctuating	**26.0%**
Declining	**19.3%**
Always inactive	**9.1%**

Trajectories of PA in adulthood, 4-class model, **men**



Always active	**40.3%**
Fluctuating/increasing	**26.7%**
Declining	**22.0%**
Always inactive	**11.0%**

Outline

- Background & Aims
- Methods
- Results
- Discussion

Key Findings

- Linguistic feature of PA, both as **active** and **spectator** possible selves, strongly associated with PA across early and mid adulthood in **male**
- In **females**, the association is **not significant** after we adjust for confounders

Future Work

- Explore the association of PA identities with trajectories of PA
- Dealing with missing data using multiple imputation

Implications

- Evidence for a PA identity that is identifiable by age 11 and helps predict adult PA
- Individuals face life course barriers to PA. Does a stronger PA identity reduce likelihood of inactivity and/or increase likelihood of returning to PA after spell of inactivity?
- Can policy foster PA identity in childhood and support it throughout adulthood?
- Promoting PA not just as a behaviour, but fostering it as part of individuals' identity may have a stronger and more persistent effect

Thank you!!

JD Carpentieri, B. Pongiglione, M. L. Kern, A. Goodman,
N. Gupta & H.A. Schwartz

Using childhood essays to predict cognitive functioning in mid life

Martina Narayanan, Alissa Goodman,
Margaret L Kern & H. Andrew Schwartz

CENTRE FOR LONGITUDINAL STUDIES

THE UNIVERSITY OF MELBOURNE

Stony Brook University

60 Years of our Lives Conference

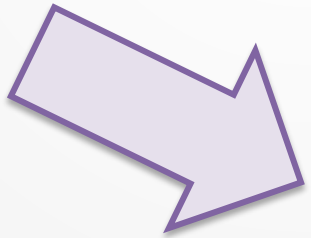
London, 9 March 2018

Background

- Can we generate scores based on the age 11 essays that capture cognitive ability?
- How do these scores correlate with cognitive tests at age 11?
- Do linguistic features based on the essays predict cognitive functioning at age 50?
- Do they do so over and above survey controls?

Outcome measures: Cognitive functioning at age 50

- Animal naming: Verbal fluency
- Letter cancellation: Attention, mental speed and visual scanning
  - Accuracy score
  - Speed score
- Memory: Verbal learning, recall and memory
  - Word list recall
  - Delayed word list recall



Principal component analysis:

Cognitive functioning at age 50

Linguistic features

- Specific feature: Complexity score
- Topics

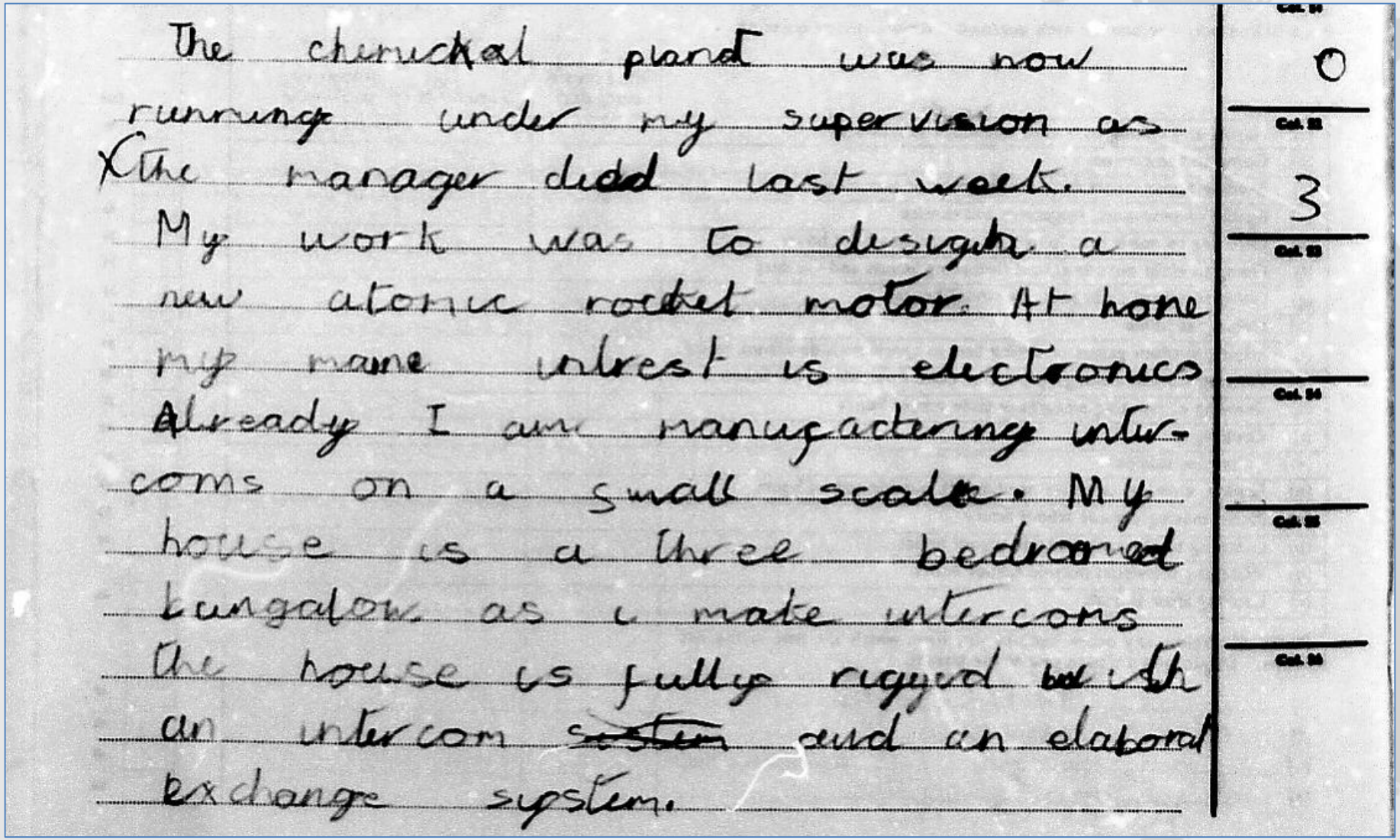
Complexity score

- How complex is the language in an essay

  = Average number of characters per word

- Scores were generated as follows:
  - → Extract all words in an essay
  - → Count the number of characters for each word
  - → Take the average
    - Correlates with general cognitive ability at age 11: r = 29***

Example Essay for High Complexity

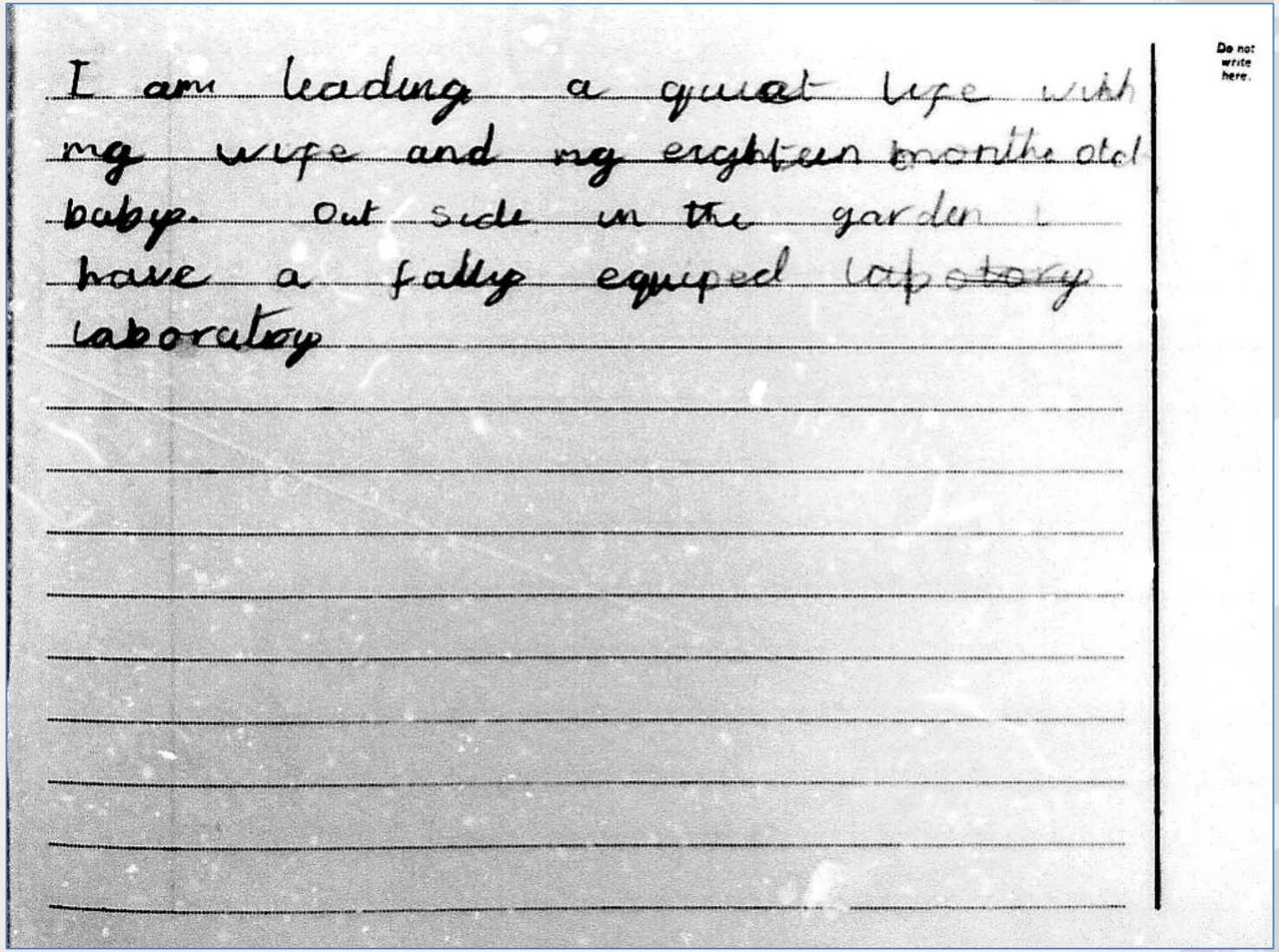


The chemickal plant was now
runnung under my supervision as
X the manager dudd last week.
My work was to design a
new atomic rocket motor. At home
my mane intrest is electronics
Already I am manufacturing inter-
coms on a small scale. My
house is a three bedroomed
bungalow as i make intercoms
the house is fully rigged with
an intercom ~~system~~ and an elabord
exchange system.

0

3

Example Essay for High Complexity



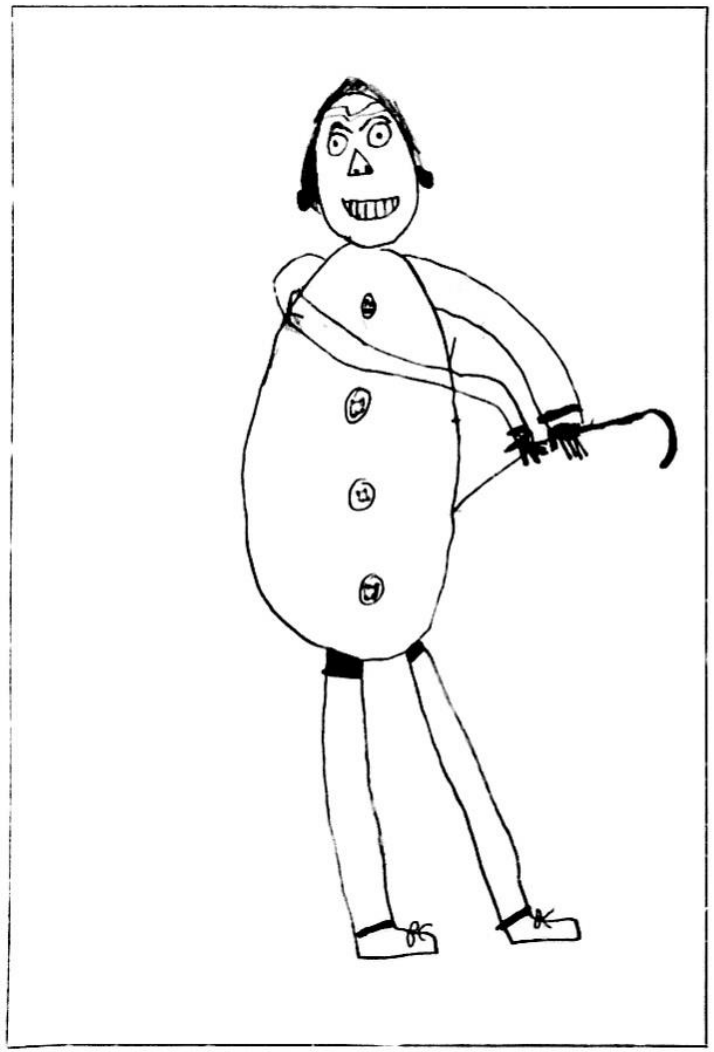
I am leading a quiet life with my wife and my eighteen months old baby. Out side in the garden I have a fully equiped laboratory

Do not write here.

Example Essay for High Complexity

“The chemical plant was now running under my supervision as the manager died last week. My work was to disign a new atomic rocket motor. At home my main intrest is electronics Already I am manufactering inturcoms on a small scale. My house is a three bedroomed bungalow. as i make intercoms the house is fully rigged with an intercom and an elaborat exchange system.@@I am leading a quiat life with my wife and my eighteen monthe old baby. Out side in the garden i have a fally equiped laboratry”

High Complexity Score: age 7 ‘draw a man’

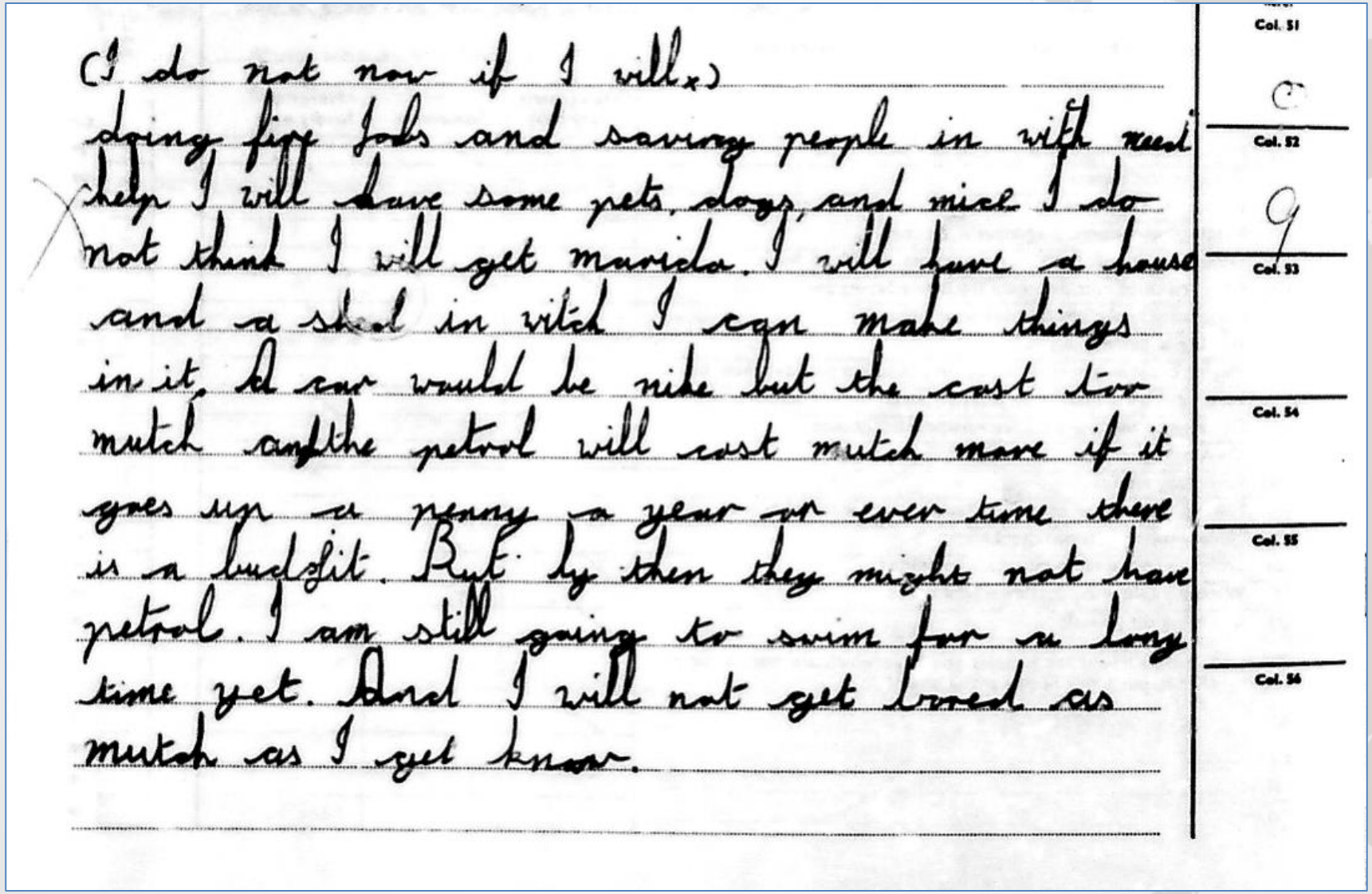


1	✓	51	
2		52	
3		52	✓
4	✓	54	✓
5	✓	55	✓
6	✓	56	
7		57	
8		58	
9	✓	59	
10	✓	60	
11	✓	61	
12	✓	62	
13		63	✓
14		64	
15		65	
16		66	
17		67	
18	✓	68	
19	✓	69	
20		70	
21	✓	71	
22	✓	72	
23		73	
24	✓	74	
25	✓	75	
26		76	
27		77	
28		78	
29		79	
30	✓	80	
31		81	
32		82	
33	✓	83	
34		84	
35	✓	85	
36		86	
37		87	
38		88	
39	✓	89	
40	✓	90	
41	✓	91	
42		92	
43	✓	93	
44	✓	94	
45		95	
46	✓	96	
47		97	
48	✓	98	
49		99	
50	✓	100	

Total Score

2	6

Example Essay for Low Complexity



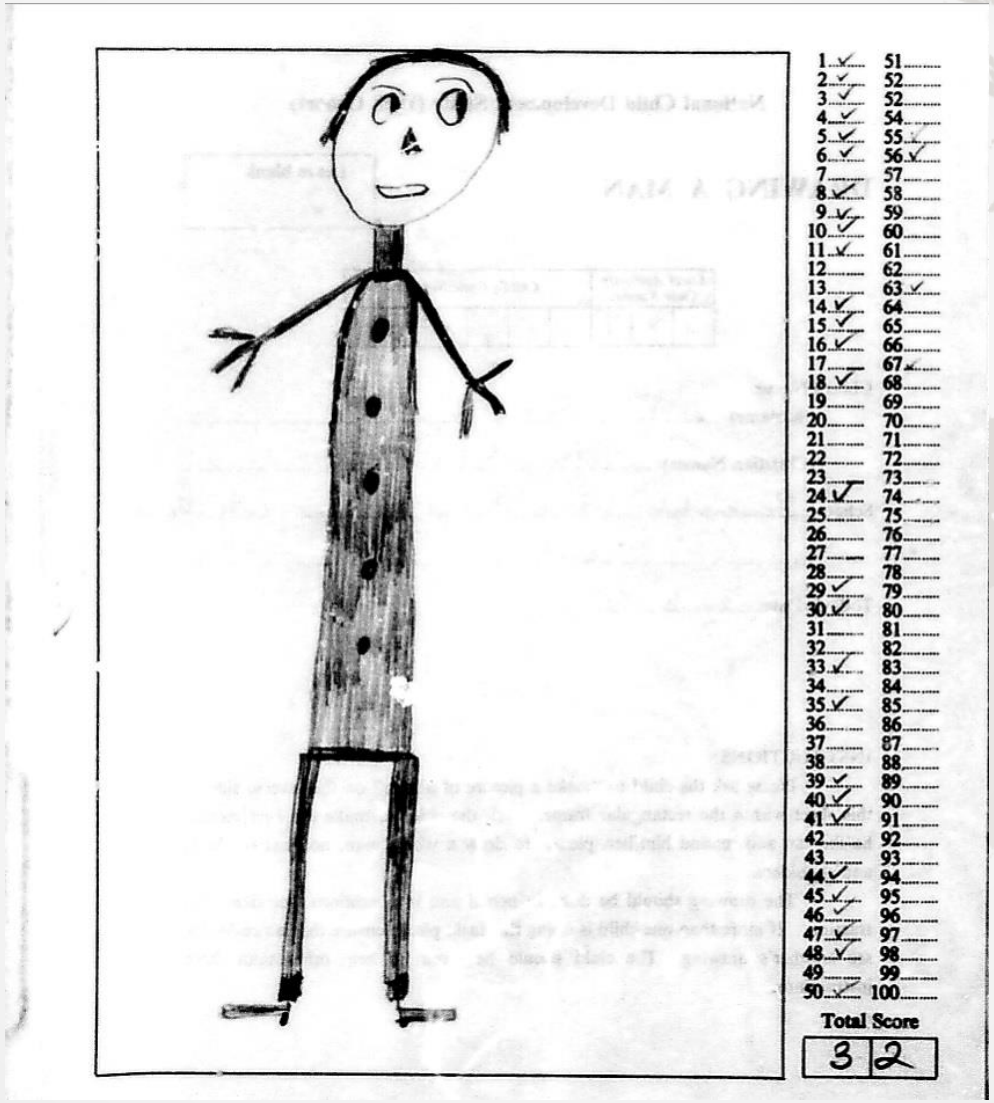
(I do not now if I will.)
doing fire fobs and saving people in with need
help I will have some pets, dogs, and mice I do
not think I will get marida. I will have a house
and a shed in with I can make things
in it. A car would be nice but the cost too
mutch and the petrol will cost mutch more if it
goes up a penny a year or ever time there
is a budget. But by then they might not have
petrol. I am still going to swim for a long
time yet. And I will not get bored as
mutch as I get know.

Col. 51
C
Col. 52
9
Col. 53
Col. 54
Col. 55
Col. 56

Example Essay for Low Complexity

“doing fire jobs and saving people in with need help I will have some pets, dogs, and mice I do not think I will get marid*, I will have a house and a shed in witch I cna make things in it. A car would e ni*e but the cast too mutch andthe petrol will cost mutch more if it goes up a penny a year or ever time there is a ******. But by then they might not have petrol . I am still going to swim for a long time yet. And I will not get bored as mutch as I get know”

Low Complexity Score: Age 7 ‘draw a man’



1	✓	51	
2	✓	52	
3	✓	52	
4	✓	54	
5	✓	55	✓
6	✓	56	✓
7		57	
8	✓	58	
9	✓	59	
10	✓	60	
11	✓	61	
12		62	
13		63	✓
14	✓	64	
15	✓	65	
16	✓	66	
17		67	✓
18	✓	68	
19		69	
20		70	
21		71	
22		72	
23		73	
24	✓	74	
25		75	
26		76	
27		77	
28		78	
29	✓	79	
30	✓	80	
31		81	
32		82	
33	✓	83	
34		84	
35	✓	85	
36		86	
37		87	
38		88	
39	✓	89	
40	✓	90	
41	✓	91	
42		92	
43		93	
44	✓	94	
45	✓	95	
46	✓	96	
47	✓	97	
48	✓	98	
49		99	
50	✓	100	

Total Score

3	2

Covariates

Minimum set:

- General cognitive ability age 11
- Gender

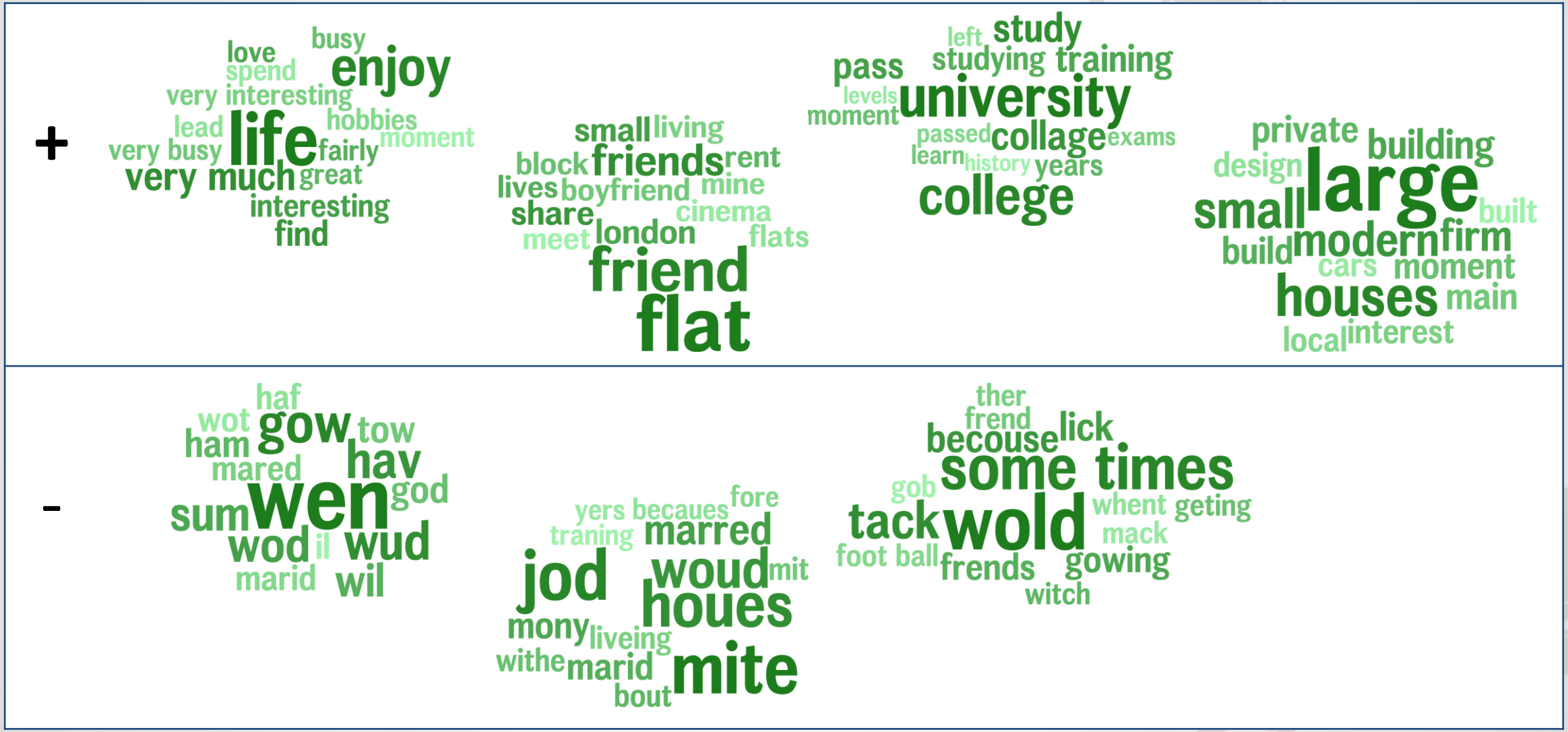
Large set:

- Psychological distress age 11
- Social class age 11
- Parent education
- Age mother
- Age father
- Region age 11
- Birthweight

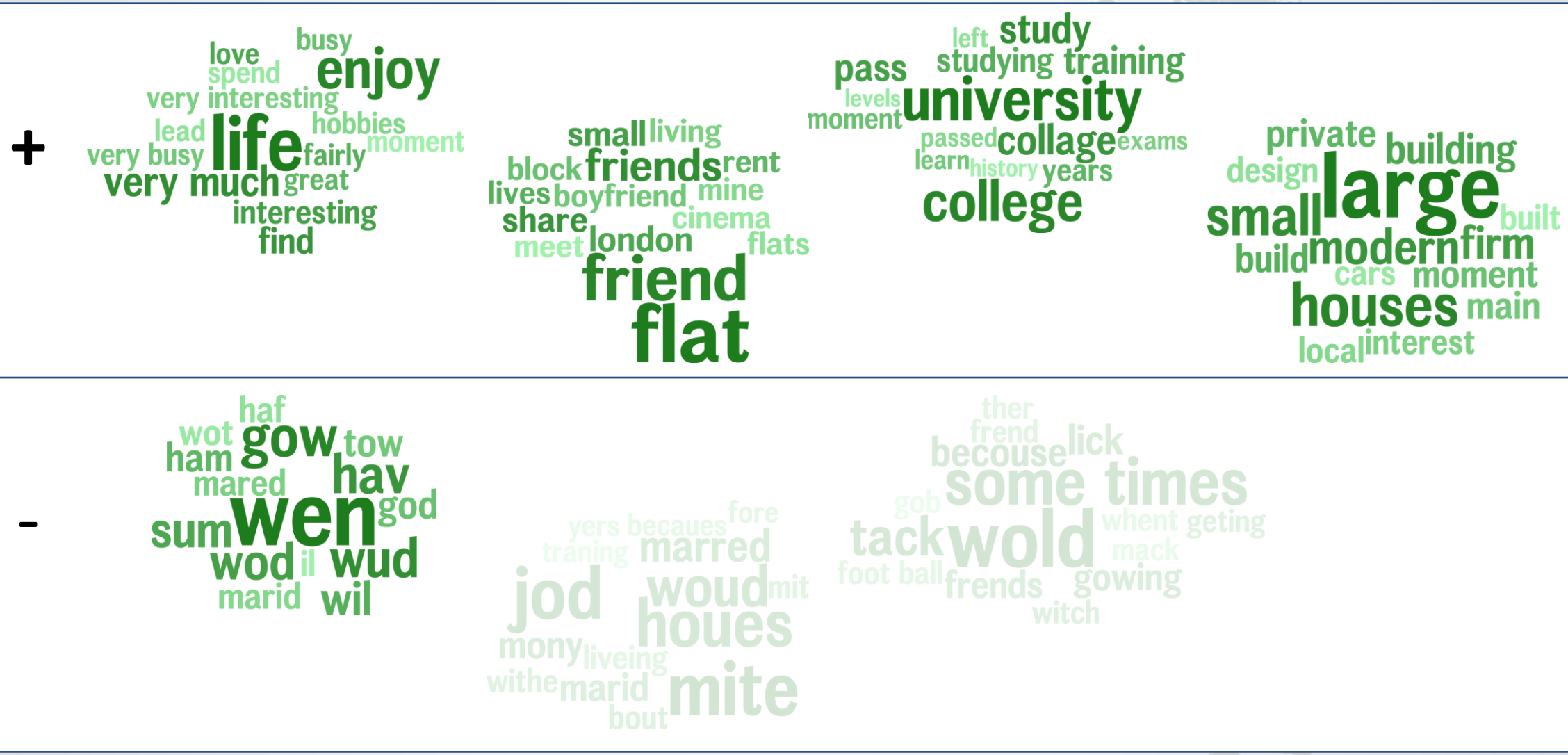
Linear regression results: Complexity score

	Cognitive functioning age 50	
1. Basic model	β	R^2
Essays complexity score	.03**	.00
2. Including cognitive tests score		.05
Essays complexity score	.04***	
Cognitive tests general ability	.23***	
3. Including set of covariates		.07
Essays complexity score	.03**	
Cognitive tests general ability	.22***	

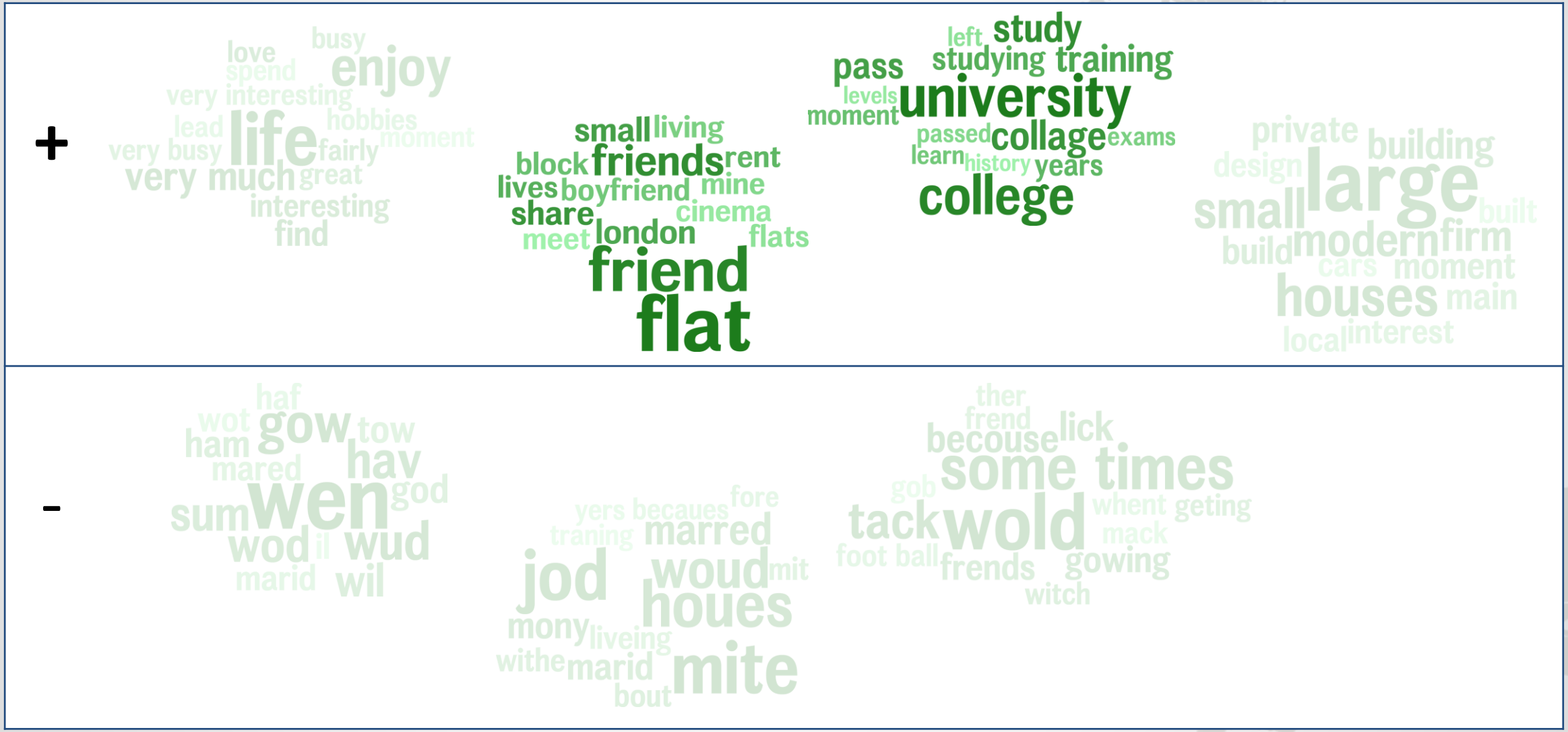
Topics Predicting Age 50 Cognitive Function



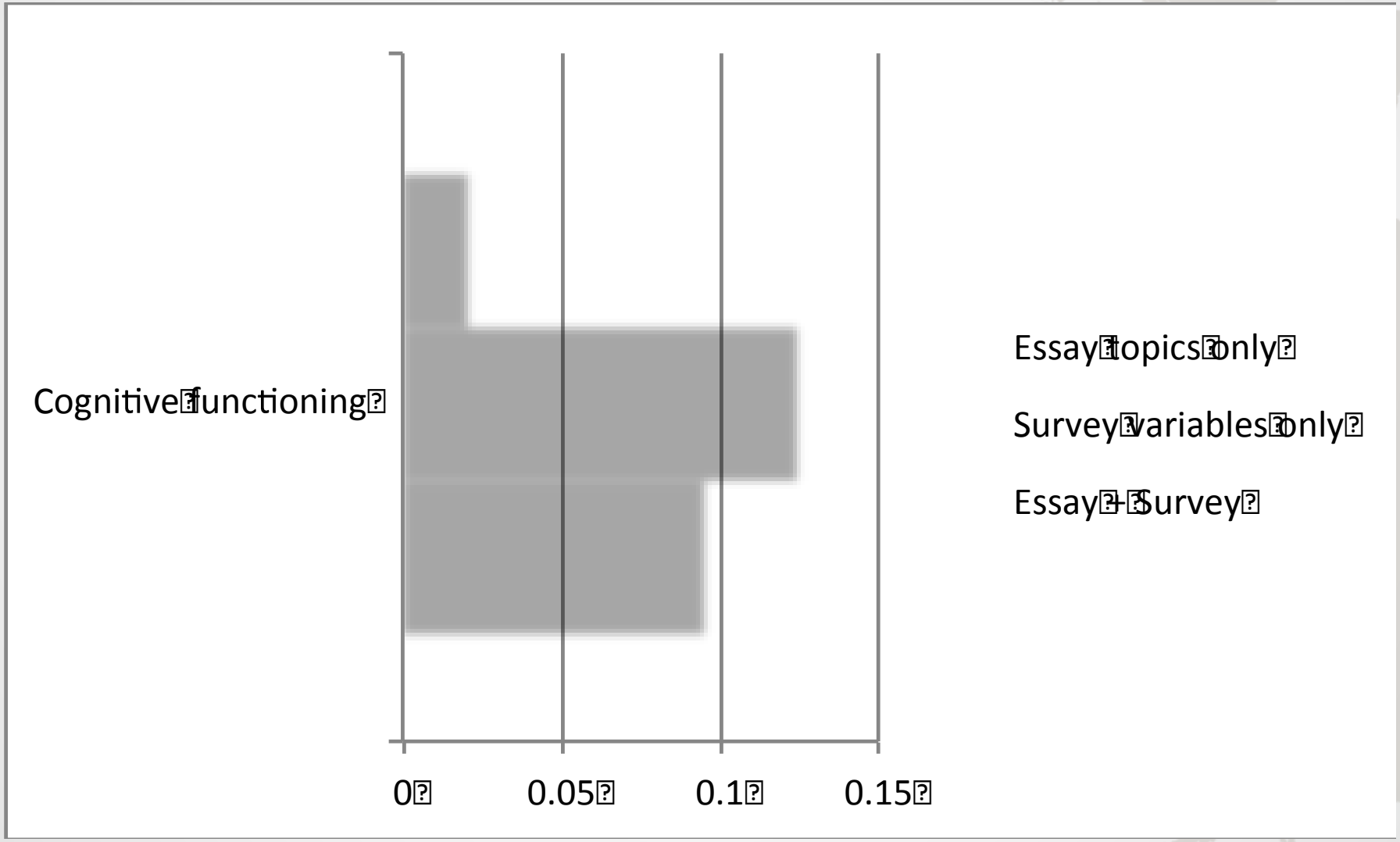
Predicting Age 50 Cognitive Function (Minimum Controls)



Predicting Age 50 Cognitive Function (All Controls)



Explained variance (R^2) for multivariate regression models



Preliminary Conclusions

- Linguistic features at age 11 can predict cognitive functioning at age 50
- But they don’t outperform survey variables

Remaining Analyses

- Ngrams
- Variable selection via machine learning algorithms

Thank you!!

Martina Narayanan, Alissa Goodman,
Margaret L Kern & H. Andrew Schwartz

Which dreams came true? Can language predict socio-economic outcomes, and social mobility?

Alissa Goodman, Martina Narayanan, Anvesh Myla,
H. Andrew Schwartz, & Margaret L Kern

CENTRE FOR LONGITUDINAL STUDIES

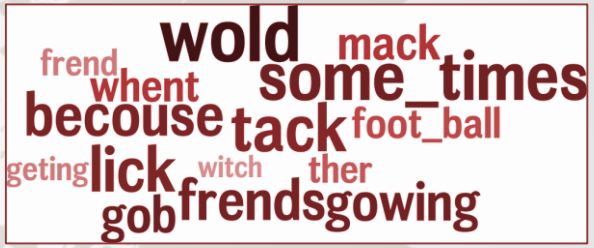
THE UNIVERSITY OF MELBOURNE

Stony Brook University

60 Years of our Lives Conference

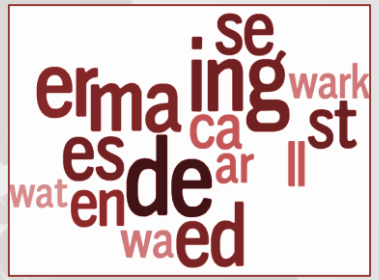
London, 9 March 2018

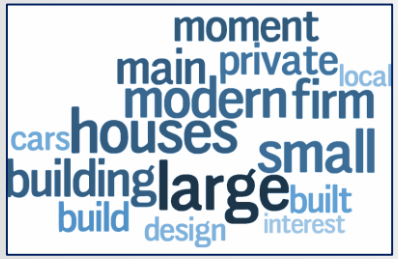
Reminder: social background very strongly shaped the essays that children wrote

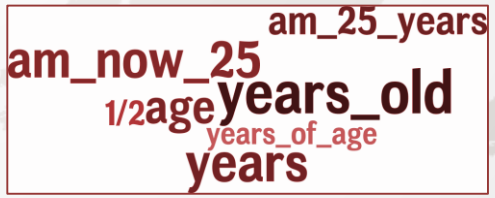


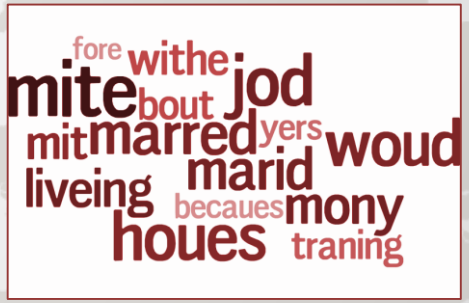
Higher social class score

Lower social class score









Social mobility project: Boy, father's social class I

I have a good job as an engineer and live on the outskirts of the town. I own a large house and equipment for astronomy which is my favourite hobby. I have a camera, cine-camera, projector, flash-gun and other equipment for use in photography. I have two dogs. Sheba which is a white boxer, and Ben which is a Red setter. I also have a collection of butterflys. In my garden I have large ponds which contain fish. I have a parrot called and two cockatoos. My parents live with me and also my brother lives five minutes walk from the house. My employer owns a wind-mill which I am allowed to visit when I please. My land consists of two and a half acres of which one acre is long grass. Counting the garden and paths and ponds and ouch I have one acre of ground which is not grass, which leaves an area of half an acre of lawn.

Social mobility project: Boy, father's social class V

I AM Now 25 years old My Job is bilder I go to Work At Seven o.clock in the morning and come home at five o.clock When I Get home in the Evening I have My dinner and the Watch T.V. and after I go to bed Then at Seven o.clock I go to Work A man comes for me at twelve o.clock I have Some thing to eat We have half an hour* the We get back to Work, from Monday to Friday I Work. I get 30xxxx a Week I like this Job Very Much. At night Some times I go Out to the picters. on a Sunday i go out to play football and on Saterday I go Swimming I like Swimming very much I can dive and Swim under water but When the Week end is over I go back to Work. In about one year time I am going to leve this Job and Joen the Army for two years. And after that I Will go to London and look for A Job and then biy a car

Transition Matrices for Income

	Poorest	2	3	4	Richest
Poorest	24.13	21.35	20.85	16.09	17.58
2	24.09	23.11	20.04	17.39	15.36
3	19.97	22.5	21.77	19.64	16.12
4	16.35	18.62	20.94	23.53	20.56
Richest	13.55	17.2	17.99	22.51	28.75
Total	**19.49**	**20.51**	**20.32**	**19.93**	**19.75**

Transition Matrices for Social Class

	I	II	IIInm	IIIm	IV & V	
I	23.8	56.76	12.39	6.48	0.56	8.35
II	13.16	51.37	22.37	9.76	3.34	21.94
IIInm	7.14	47.7	27.48	15.01	2.66	10.04
IIIm	5.73	36.81	25.41	25.6	6.45	49.85
IV & V	4.57	32.01	23.17	27.13	13.11	9.82
	9.2	42.84	23.53	19.1	5.34	

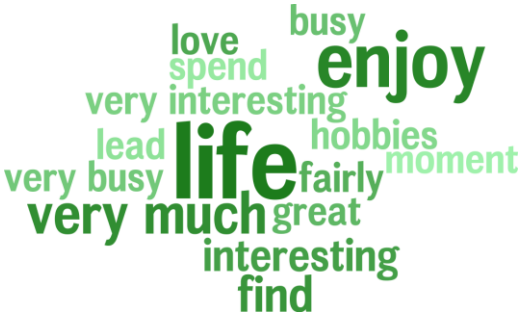
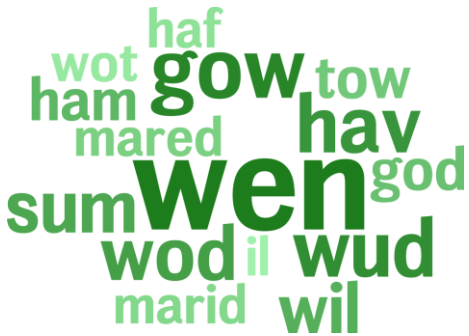
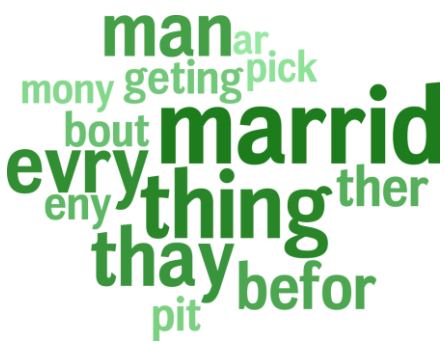
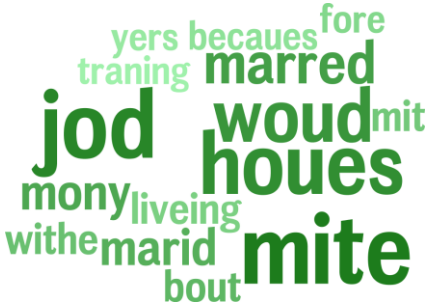
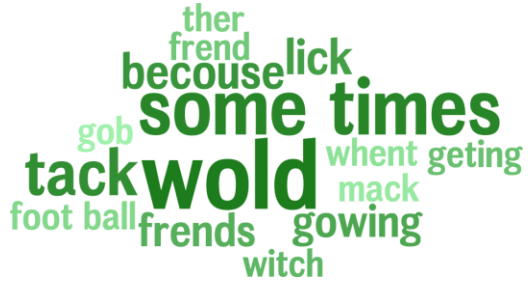
Parent Class

Cohort Member Class

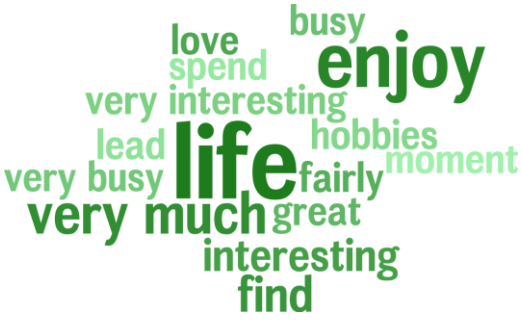
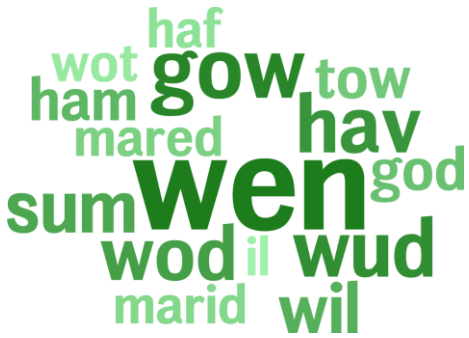
Does language predict socio-economic outcomes, and social mobility?

1. Which topics are individually predictive of future socio-economic (SES) outcomes?
    + Control for parental SES (social mobility model)
    + A wider set of survey controls
2. How much can these topics jointly predict future socio-economic (SES) - (machine learning models, trained and then predicted out of sample?)
    + Control for parental SES (social mobility model)

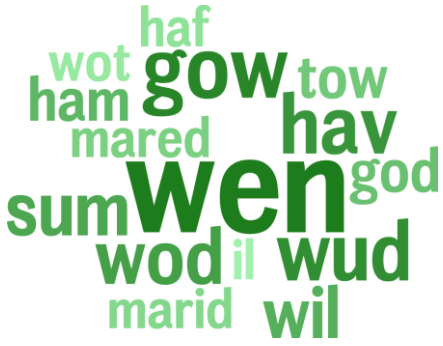
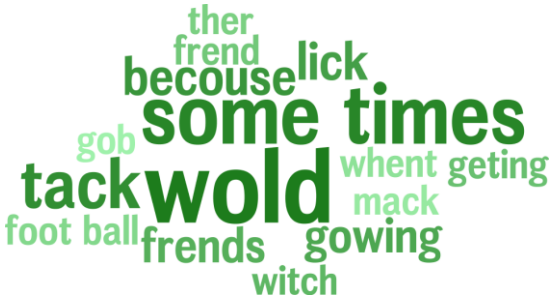
Topics predictive of income in men

Higher				
Lower				

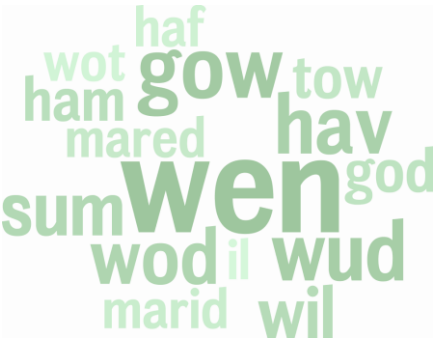
Topics predictive of income in men

Higher				
Lower				

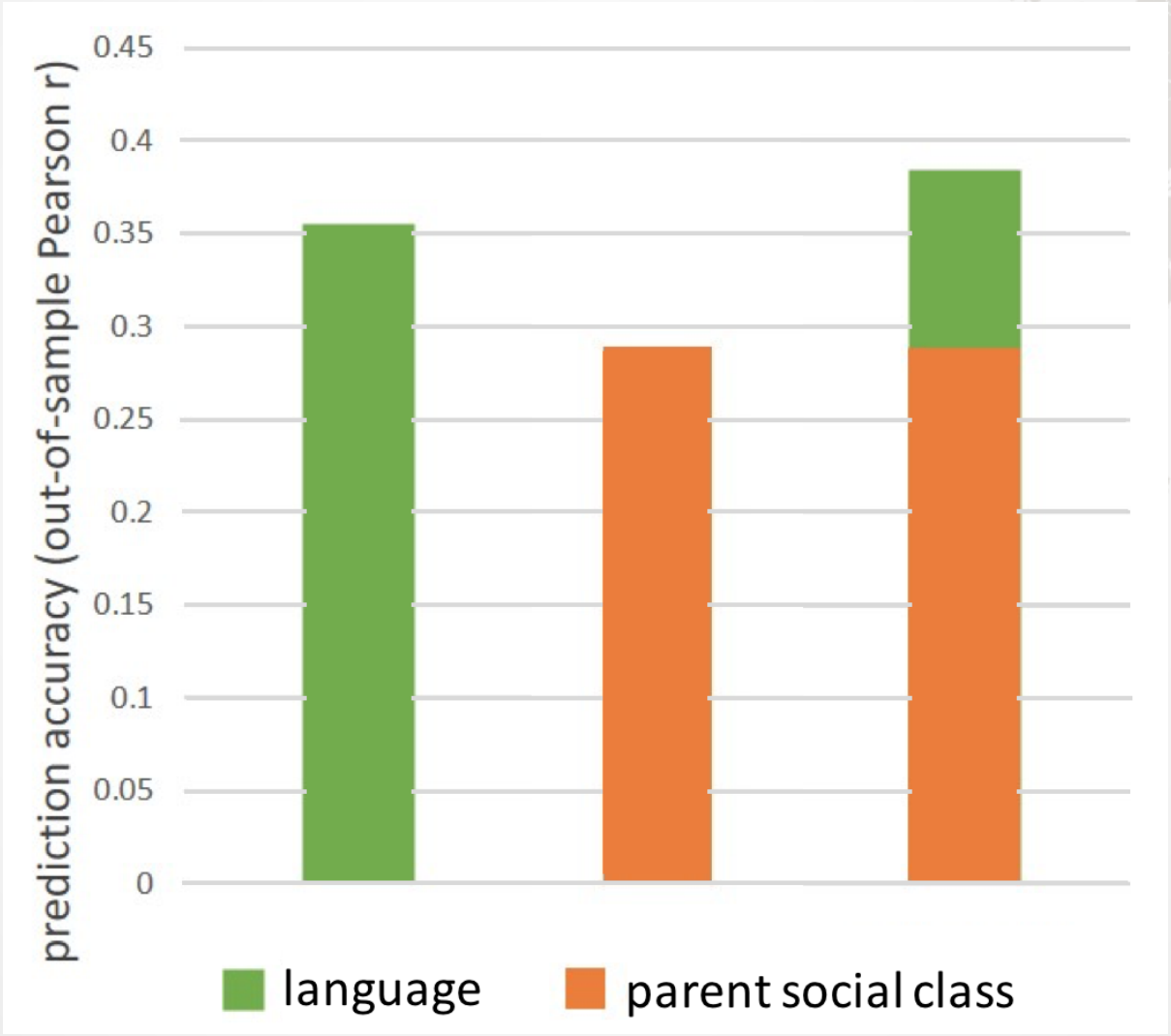
Topics predictive of income in women

Higher				
Lower				

Topics predictive of income in women

Higher				
Lower				

Social Class Predictive Accuracy



Future Work

- Refinement of our models: Interacted models (resilience; upward mobility); controls
- Other linguistic features: Which dreams came true? (own and partner occupations); Complexity, affect; use of particular words

Thank you!!

Alissa Goodman, Martina Narayanan,
H. Andrew Schwartz, & Margaret L Kern

Questions and Discussion

Concluding Thoughts from the Panel

- Fairly unlikely that the language (written once at school) can ‘beat the survey’ … but this is probably not a fair test
- The surveys can be informative by showing which aspects of the survey this language does and doesn’t capture
- Opportunity to demonstrate predictive power into the future – most language/ outcome analysis is short-term or near future orientated
- But we may still find within the language undiscovered traits that are not covered in the surveys

Shared Task 2018

CLPsych-2018 Shared Task: Call for Participation

Predicting Current and Future Psychological Health from Childhood Essays

We invite participants for the 2018 CLPsych Shared Task.

Motivation and Background

This shared task seeks to encourage new methods not only for analyzing current language use as a signal for mental health, as in previous CLPsych shared tasks, but also for understanding childhood language as a marker of future psychological health over individual lifetimes. Predicting well-being from language in the short term is valuable, such as in improving intake assessment and monitoring. However, predictions about the long-term future, an area with little work thus far from the NLP community, can aid with another class of applications: the understanding of early life markers and development of preventative care.

The unprecedented data for this task comes from the National Child Development Study, (NCDS), also known as the 1958 British Birth Cohort Study, which follows a cohort of all children born in a single week in Great Britain in March

Follow us on Twitter

Thank you!!

Alissa Goodman, Margaret L Kern, Benedetta Pongiglione, Martina Narayanan, H. Andrew Schwartz, JD Carpentieri

Appendix

Hoped For Self: Professional footballer

‘[….] before i was a footballer, i was a brick layer, until i was twenty three we had a team in the brick layers yard, and one day when i was playing for them a man came up to me and asked me if i would like to play for manchester united of course i said yes. At that time i was smoking and found it hard to give up but in the end i managed to give up you cant smoke in football you see, so i had to give it up. [….] I have played in the F.A. cup final once but have lost one nil [….]’ (Male, N21094U)

Continuity of PA over time, Hoped For Self

‘I have got a bungalow by the sea-side. I like it by the sea because I can swim in it, and there is lots of fresh air. I would like to be a champion diver because I can dive. I would go out a lot. I would take my wife out a lot, But I wont take her out if she doesn't want to go out.’ (Male, N24157E)

Results: PA as part of projected daily routine

‘[….] my Job is a hair dresser. [….] I work from Eight o’clock in the morning to half past Eight at night. I have half day on Thursday. When I go home at night I go swimming. for half an hour.[….] I am very happy with My work and my hobbies. and I would never change With any body Else.’ (Female, N24256G)

Intergenerational continuity of PA

‘[….] when it is a hour to go to work I go out to play a game with my son Royston he is fond of football. He is in goal with the wind blowing toward me and I am in goal when it is blowing toward him [….]’ (Male, N24781Y)

‘[….] I like swimming and I take my children swimming twice a week [….]’ (Female, N24458Q)

Physical Activity Cumulative Scores Across Adulthood

Cumulative score	Distribution	
	Sample 1 (N=7,673)	Sample 2 (N=12,763)
0	12.54	12.45
1	22.74	23.12
2	22.99	23.25
3	22.25	21.99
4	19.48	19.19