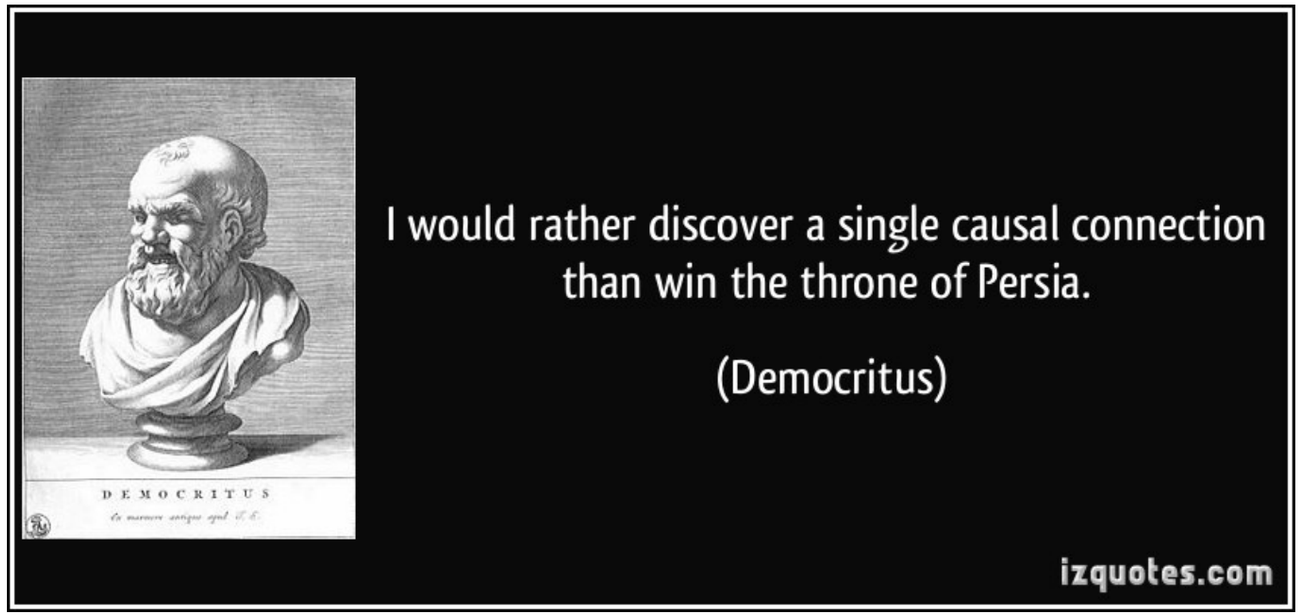
What We **Write** About

When We Write about **Causality**

T.McAndrew

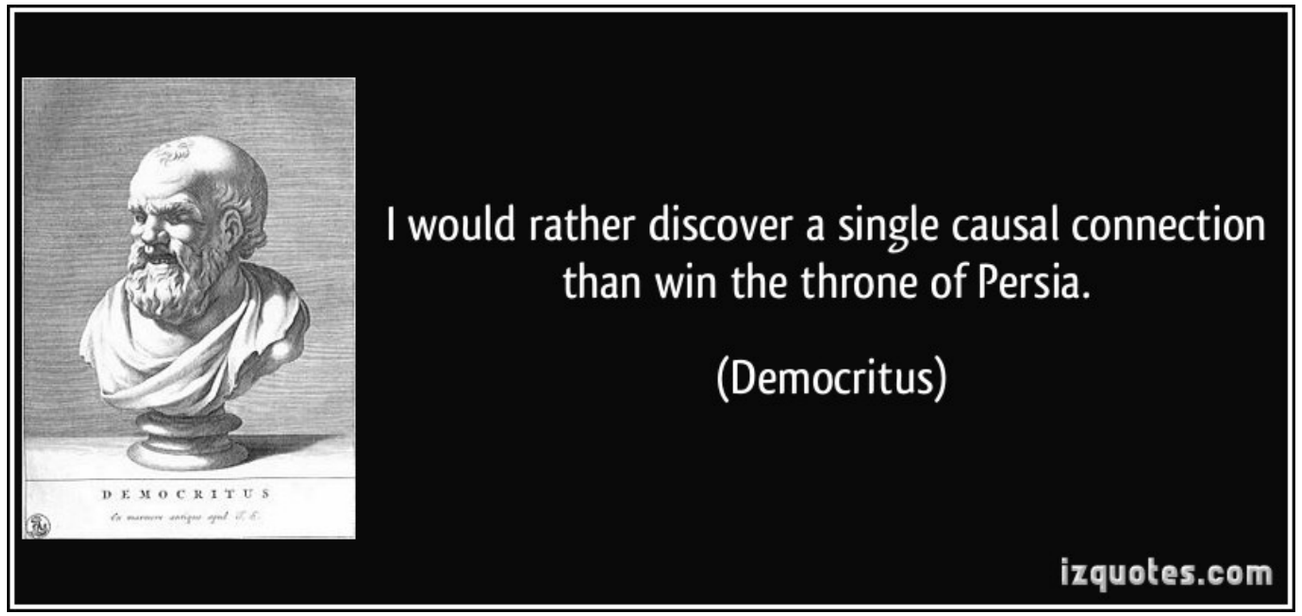
The University of Vermont

Motivation



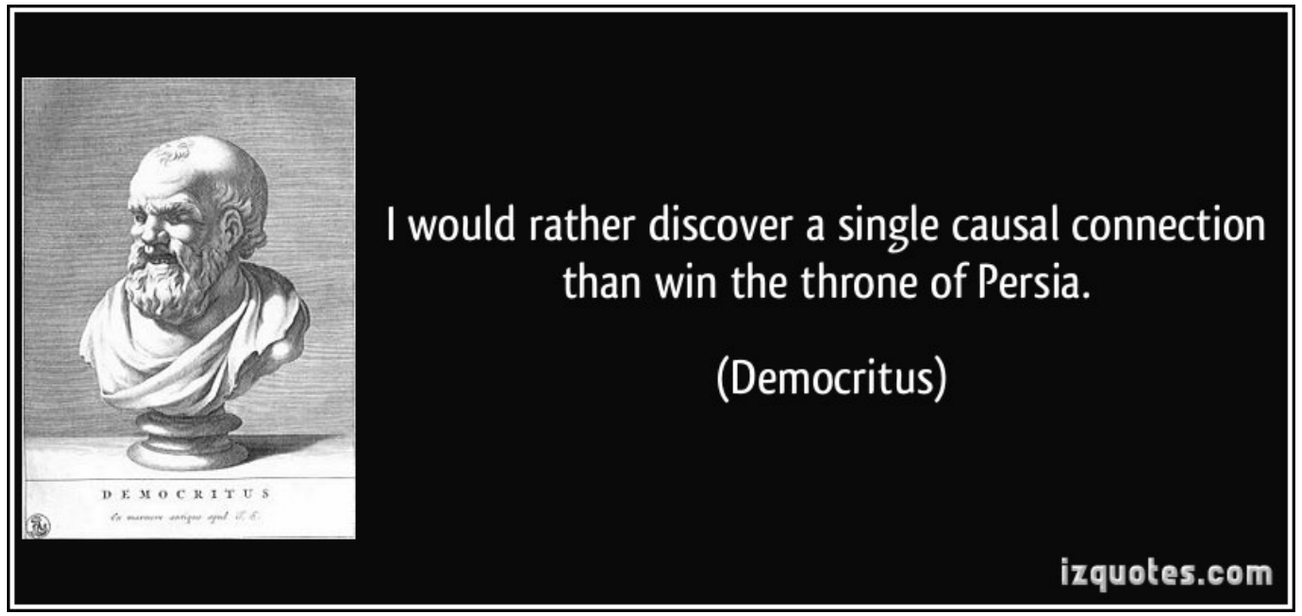
vs.

Motivation



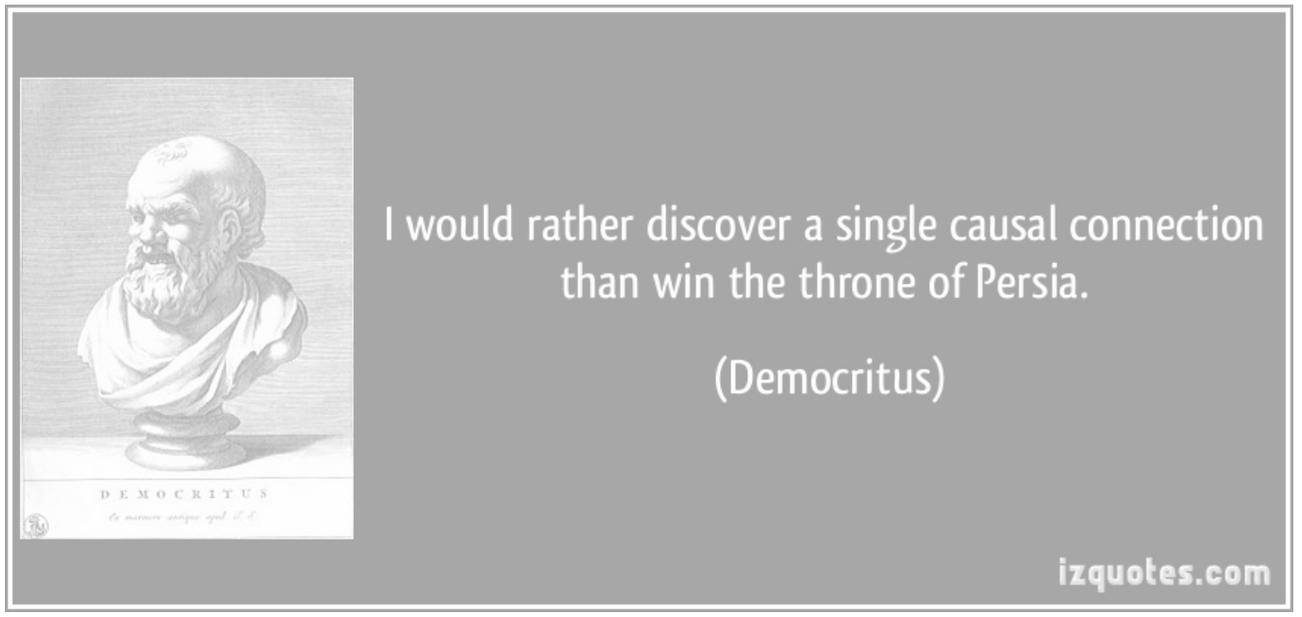
vs.

Motivation



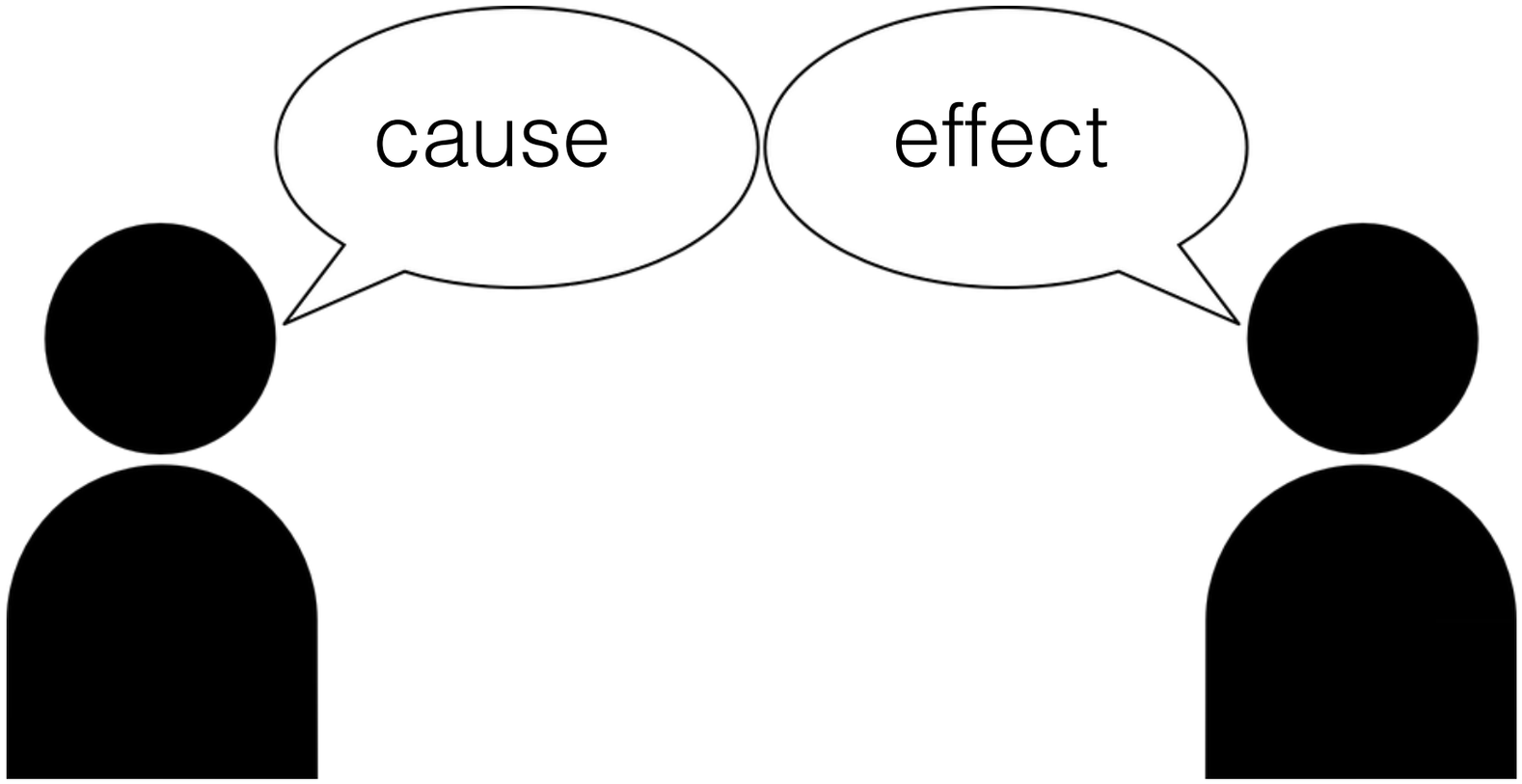
vs.

Motivation



VS.

Motivation



Methods

data

Collected

10 % Random Sample

2013

English

No Retweets

Methods

data

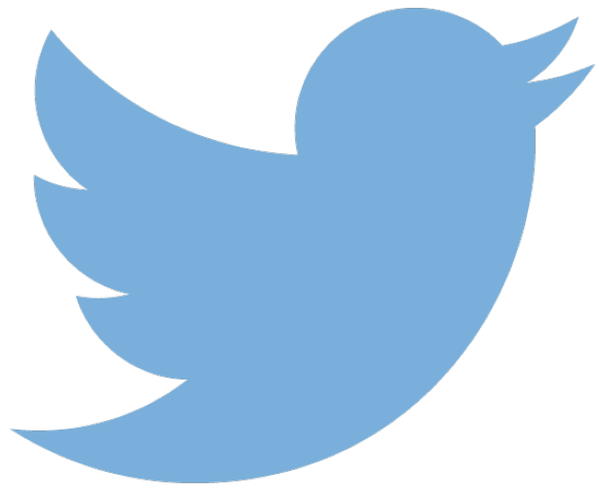
Collected

10 % Random Sample

2013

English

No Retweets



Removed

XML, #, @, punctuation

No Stemming

lower case

Methods

data

Obtaining Causal and Control Corpora

```
for d ∈ All Documents
    words = split_on_whitespace(d)
    if (Cause,Caused,Causes) ∈ words :
        Add to Causal Corpus
    else:
        Add to Control Corpus
```

Match Control documents on

Methods -(1/3)

Comparing Corpora

$$OR(x) = \frac{p_C(x)/(1 - p_C(x))}{p_N(x)/(1 - p_N(x))}$$

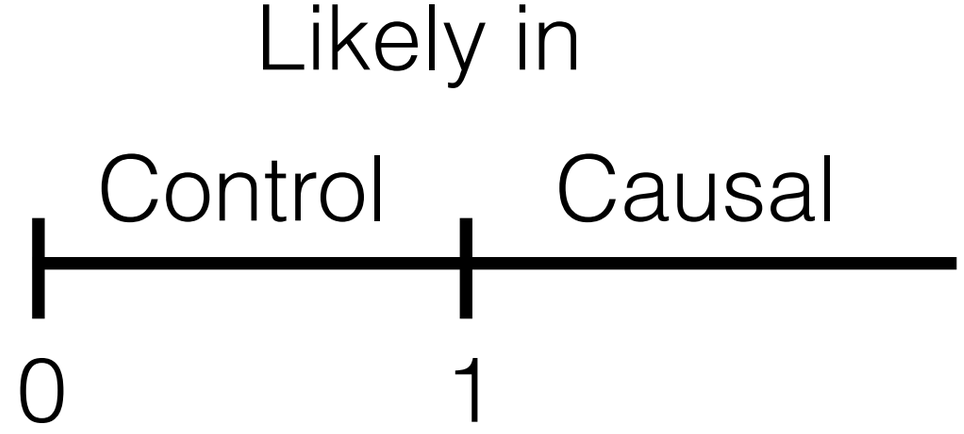
where:

$$p(x) = f(x) \Big/ \sum_{x' \in V} f(x')$$

Methods -(1/3)

Comparing Corpora

$$OR(x) = \frac{p_C(x)/(1-p_C(x))}{p_N(x)/(1-p_N(x))}$$



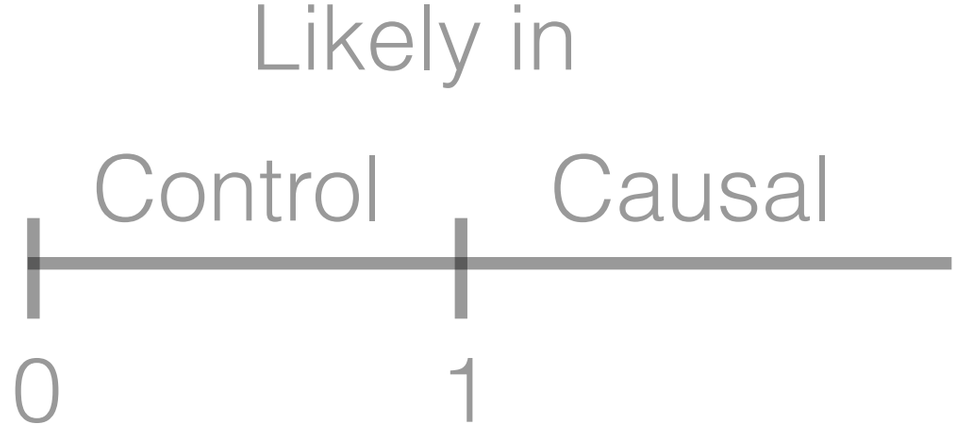
where:

$$p(x) = f(x) \Big/ \sum_{x' \in V} f(x')$$

Methods -(1/3)

Comparing Corpora

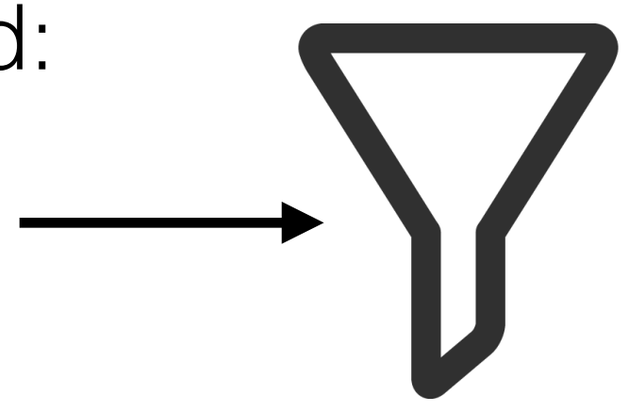
$$OR(x) = \frac{p_C(x)/(1-p_C(x))}{p_N(x)/(1-p_N(x))}$$



where:

$$p(x) = f(x) \Big/ \sum_{x' \in V} f(x')$$

and:

w  tf-idf$(w) >$ 90th percentile

$$\text{tf-idf}(w) = \log f(w) \times \log\left(\frac{\text{D}}{df(w)}\right)$$

Methods - (2/3)
Sentiment Analysis

MT

Goal: $g(w) \rightarrow \text{sentiment}$

Composite top 5,000 (freq.)

10,222 Words

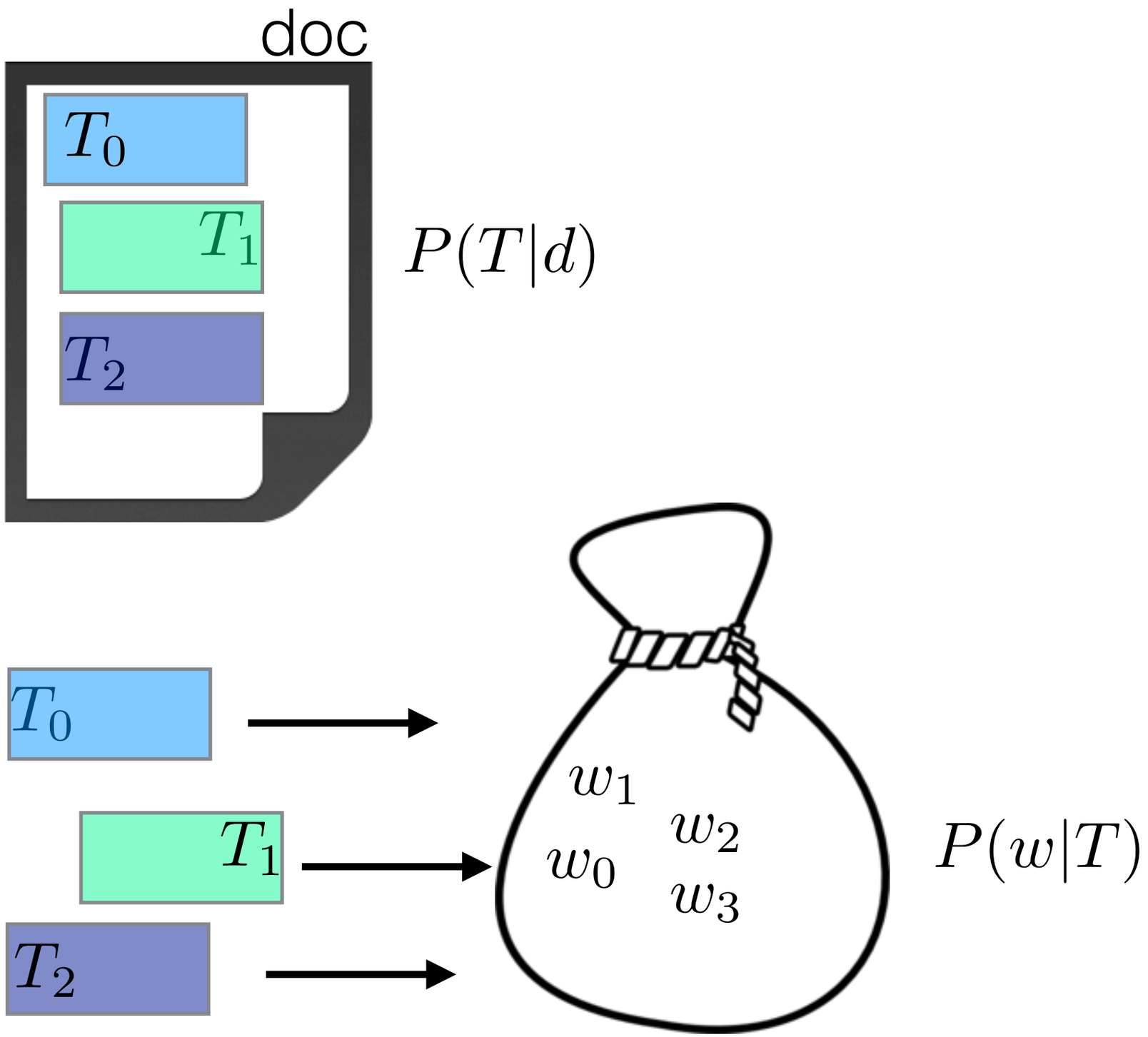
Corpus Level:

$$\sum_{w \in \bar{V}} f(w)s(w) \Big/ \sum_{w' \in \bar{V}} f(w')$$

Methods

Topic Modeling

LDA



Results

Comparing Corpora

i$_{NNS}$ actually$_{RB}$ hate$_{VBP}$ drama$_{NN}$ it$_{PRP}$ causes$_{VBZ}$ so$_{RB}$ much unnecessary$_{JJ}$ stress$_{NN}$

the$_{DT}$ bieber$_{NNP}$ family$_{NN}$ aka$_{VBZ}$ the$_{DT}$ cutest$_{JJS}$ family

this$_{DT}$ one$_{NN}$ problem$_{NN}$ has$_{VBZ}$ caused$_{VBN}$ so$_{RB}$ much$_{JJS}$ hurt$_{NN}$ and$_{CC}$ pain$_{NN}$ [...]

freezing$_{VBG}$ rain$_{NN}$ causes$_{NNS}$ thousands$_{NNS}$ to$_{TO}$ lose$_{VB}$ power$_{NN}$ across$_{IN}$ southern$_{JJ}$ ontario$_{NN}$

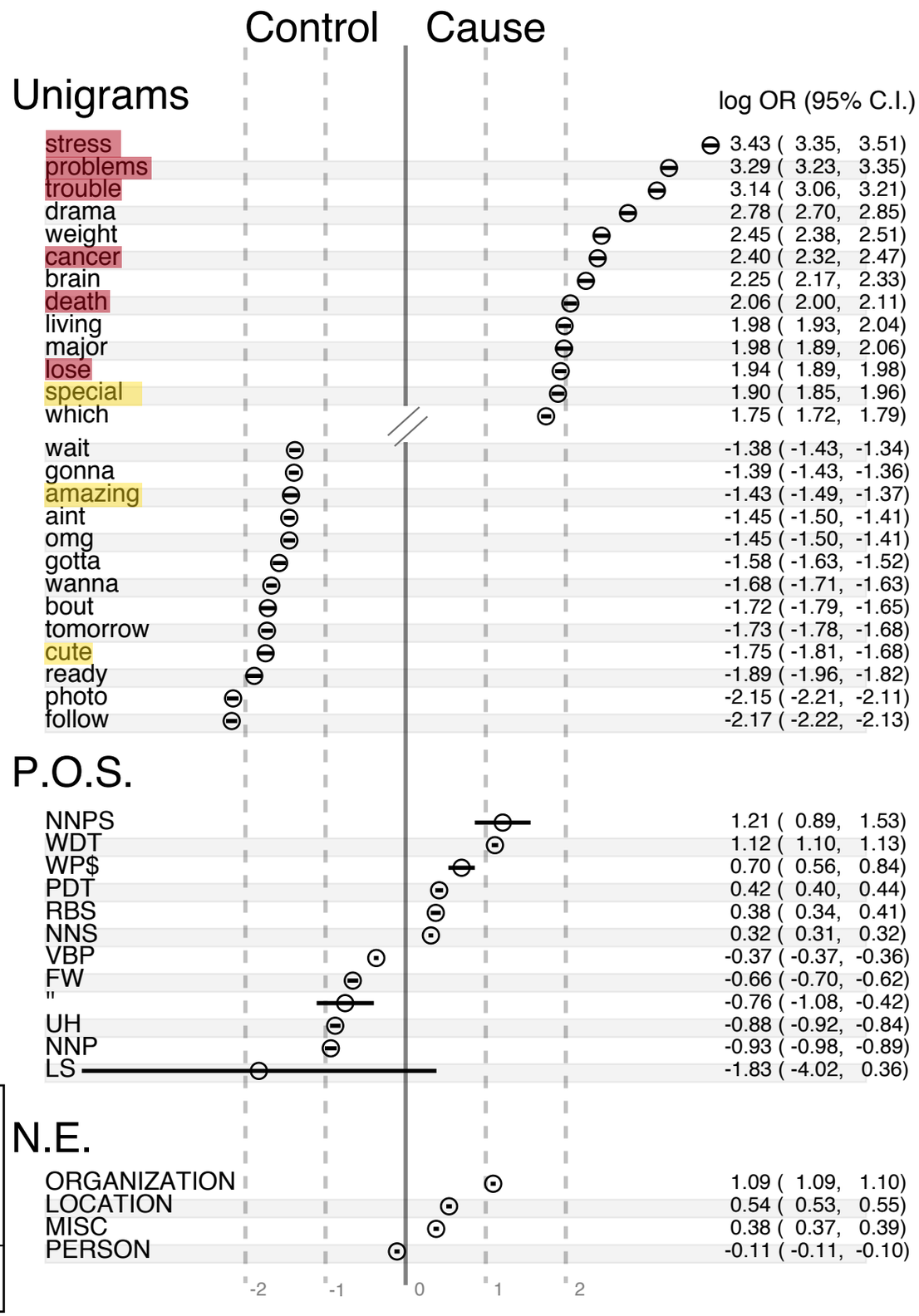
he$_{PRP}$ gon$_{MD}$ play$_{VB}$ until$_{IN}$ he$_{PRP}$ wins$_{VBZ}$ more$_{RBR}$ or$_{CC}$ until$_{IN}$ he$_{PRP}$ cant$_{VB}$ nomo$_{JJ}$ [...]

i$_{NN}$ think$_{VBP}$ the$_{DT}$ ps$_{NN}$ is$_{VBZ}$ a$_{DT}$ amazing$_{JJ}$ product$_{NN}$ its$_{PRP\$}$ worth$_{JJ}$ buying$_{NN}$

londons$_{O}$ appolo$_{O}$ theatre$_{O}$ collapses causing injuries eyewitnesses have described the chaos and panic

kevin$_{P}$ we have had tons of snow in bowermanville we could help the slopes at blue$_{L}$ mountain$_{L}$

Part of Speech (P.O.S)			**Named Entities (N.E.)**
WDT/ WP: Wh-determiner/pronoun			O:Organization
(N/NN)+(S/P): Noun plural/proper			L:Location
(P)+DT/CC:(pre)deter./conjunc.			P:Person
JJ+(S): adjective/superlative			M:Miscellaneous
MD/FW/LS:Modal/Foreign Word/List Item	VB+(GPNZ): Verb	RB+(RS): adverbs	
(IN,PRP+$)/to: preposition (possess)/to	UH:Interjection	":quotes	



Results

Comparing Corpora

i$_{NNS}$ actually$_{RB}$ hate$_{VBP}$ drama$_{NN}$ it$_{PRP}$ causes$_{VBZ}$ so$_{RB}$ much unnecessary$_{JJ}$ stress$_{NN}$

the$_{DT}$ bieber$_{NNP}$ family$_{NN}$ aka$_{VBZ}$ the$_{DT}$ c[illegible]$_{JJS}$ family

this$_{DT}$ one$_{NN}$ problem$_{NN}$ has$_{VBZ}$ caused$_{VBN}$ so$_{RB}$ much$_{JJS}$ hurt$_{NN}$ and$_{CC}$ pain$_{NN}$ [...]

freezing$_{VBG}$ rain$_{NN}$ causes$_{NNS}$ thousands$_{NNS}$ to$_{TO}$ lose$_{VB}$ power$_{NN}$ across$_{IN}$ southern$_{JJ}$ ontario$_{NN}$

he$_{PRP}$ gon$_{MD}$ play$_{VB}$ until$_{IN}$ he$_{PRP}$ wins$_{VBZ}$ more$_{RBR}$ or$_{CC}$ until$_{IN}$ he$_{PRP}$ cant$_{VB}$ nom[illegible]$_{JJ}$ [...]

[illegible]$_{NN}$ think$_{VBP}$ the$_{DT}$ ps$_{NN}$ is$_{VBZ}$ a$_{DT}$ [illegible]$_{JJ}$ product$_{NN}$ its$_{PRP\$}$ worth$_{JJ}$ buying$_{NN}$

londons$_{O}$ appolo$_{O}$ theatre$_{O}$ collapses causing injuries eyewitnesses have described the chaos and panic

kevin$_{P}$ we have had tons of snow in bowermanville we could help the slopes at blue$_{L}$ mountain$_{L}$

Part of Speech (P.O.S)

WDT/ WP: Wh-determiner/pronoun

(N/NN)+(S/P): Noun plural/proper

(P)+DT/CC:(pre)deter./conjunc.

JJ+(S): adjective/superlative

MD/FW/LS:Modal/Foreign Word/List Item

(IN,PRP+$)/to: preposition (possess)/to

VB+(GPNZ): Verb

UH:Interjection

RB+(RS): adverbs

Named Entities (N.E.)

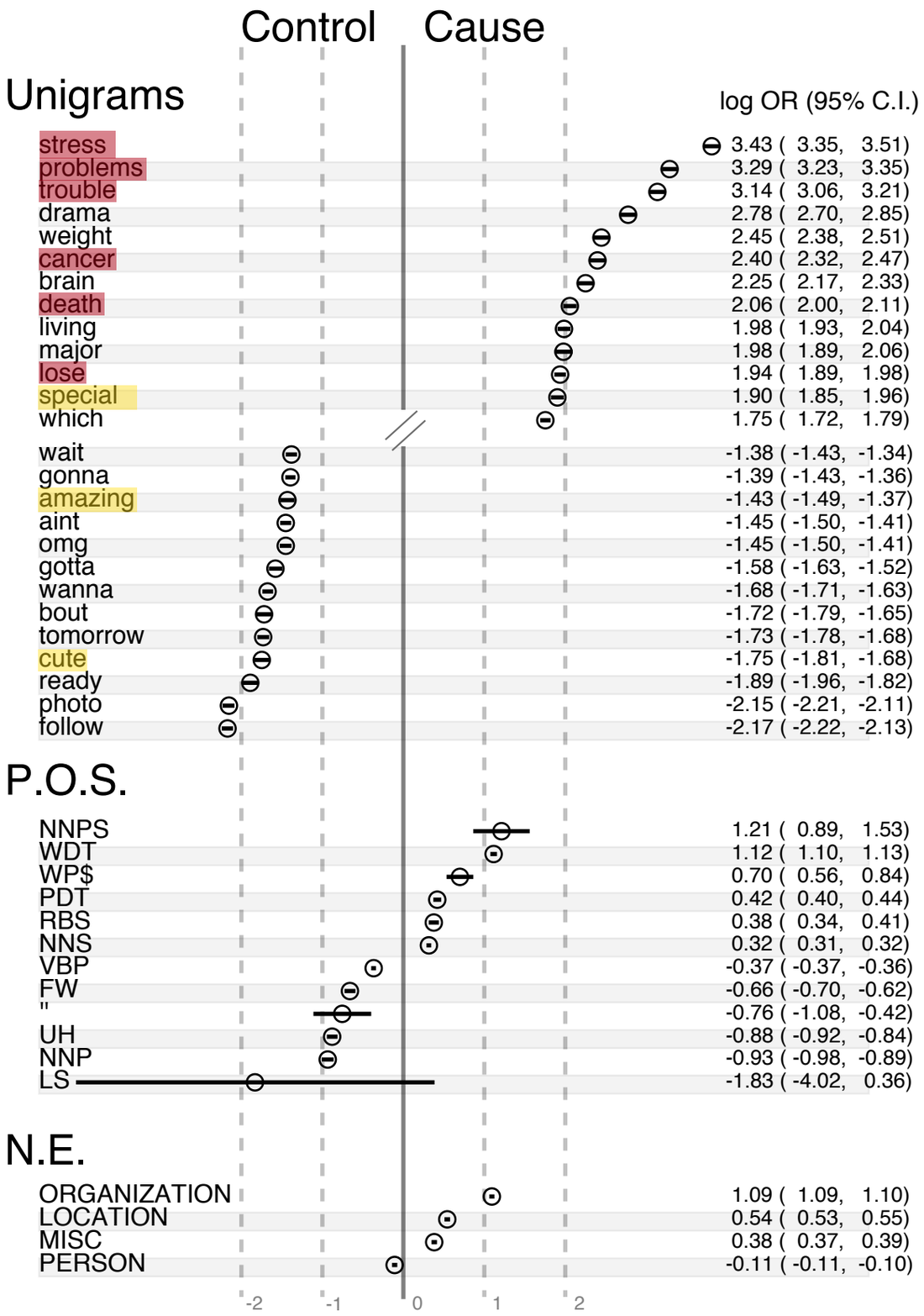
O:Organization

L:Location

P:Person

M:Miscellaneous

'quotes



Results

Anecdotal text

i$_{NNS}$ actually$_{RB}$ hate$_{VBP}$ drama$_{NN}$ it$_{PRP}$ causes$_{VBZ}$ so$_{RB}$ much unnecessary$_{JJ}$ stress$_{NN}$

the$_{DT}$ bieber$_{NNP}$ family$_{NN}$ aka$_{VBZ}$ the$_{DT}$ cutest$_{JJS}$ family

this$_{DT}$ one$_{NN}$ problem$_{NN}$ has$_{VBZ}$ caused$_{VBN}$ so$_{RB}$ much$_{JJS}$ hurt$_{NN}$ and$_{CC}$ pain$_{NN}$ [...]

freezing$_{VBG}$ rain$_{NN}$ causes$_{NNS}$ thousands$_{NNS}$ to$_{TO}$ lose$_{VB}$ power$_{NN}$ across$_{IN}$ southern$_{JJ}$ ontario$_{NN}$

he$_{PRP}$ gon$_{MD}$ play$_{VB}$ until$_{IN}$ he$_{PRP}$ wins$_{VBZ}$ more$_{RBR}$ or$_{CC}$ until$_{IN}$ he$_{PRP}$ cant$_{VB}$ nomo$_{JJ}$ [...]

i$_{NN}$ think$_{VBP}$ the$_{DT}$ ps$_{NN}$ is$_{VBZ}$ a$_{DT}$ amazing$_{JJ}$ product$_{NN}$ its$_{PRP\$}$ worth$_{JJ}$ buying$_{NN}$

londons$_{O}$ appolo$_{O}$ theatre$_{O}$ collapses causing injuries eyewitnesses have described the chaos and panic

kevin$_{P}$ we have had tons of snow in bowermanville we could help the slopes at blue$_{L}$ mountain$_{L}$

Part of Speech (P.O.S)			**Named Entities (N.E.)**
WDT/ WP: Wh-determiner/pronoun			O:Organization
(N/NN)+(S/P): Noun plural/proper			L:Location
(P)+DT/CC:(pre)deter./conjunc.			P:Person
JJ+(S): adjective/superlative			M:Miscellaneous
MD/FW/LS:Modal/Foreign Word/List Item	VB+(GPNZ): Verb	RB+(RS): adverbs	
(IN,PRP+$)/to: preposition (possess)/to	UH:Interjection	":quotes	

i$_{NNS}$ actually$_{RB}$ hate$_{VBP}$ drama$_{NN}$ it$_{PRP}$ causes$_{VBZ}$ so$_{RB}$ much unnecessary$_{JJ}$ stress$_{NN}$

the$_{DT}$ bieber$_{NNP}$ family$_{NN}$ aka$_{VBZ}$ the$_{DT}$ cutest$_{JJS}$ family

this$_{DT}$ one$_{NN}$ problem$_{NN}$ has$_{VBZ}$ caused$_{VBN}$ so$_{RB}$ much$_{JJS}$ hurt$_{NN}$ and$_{CC}$ pain$_{NN}$ [...]

freezing$_{VBG}$ rain$_{NN}$ causes$_{NNS}$ thousands$_{NNS}$ to$_{TO}$ lose$_{VB}$ power$_{NN}$ across$_{IN}$ southern$_{JJ}$ ontario$_{NN}$

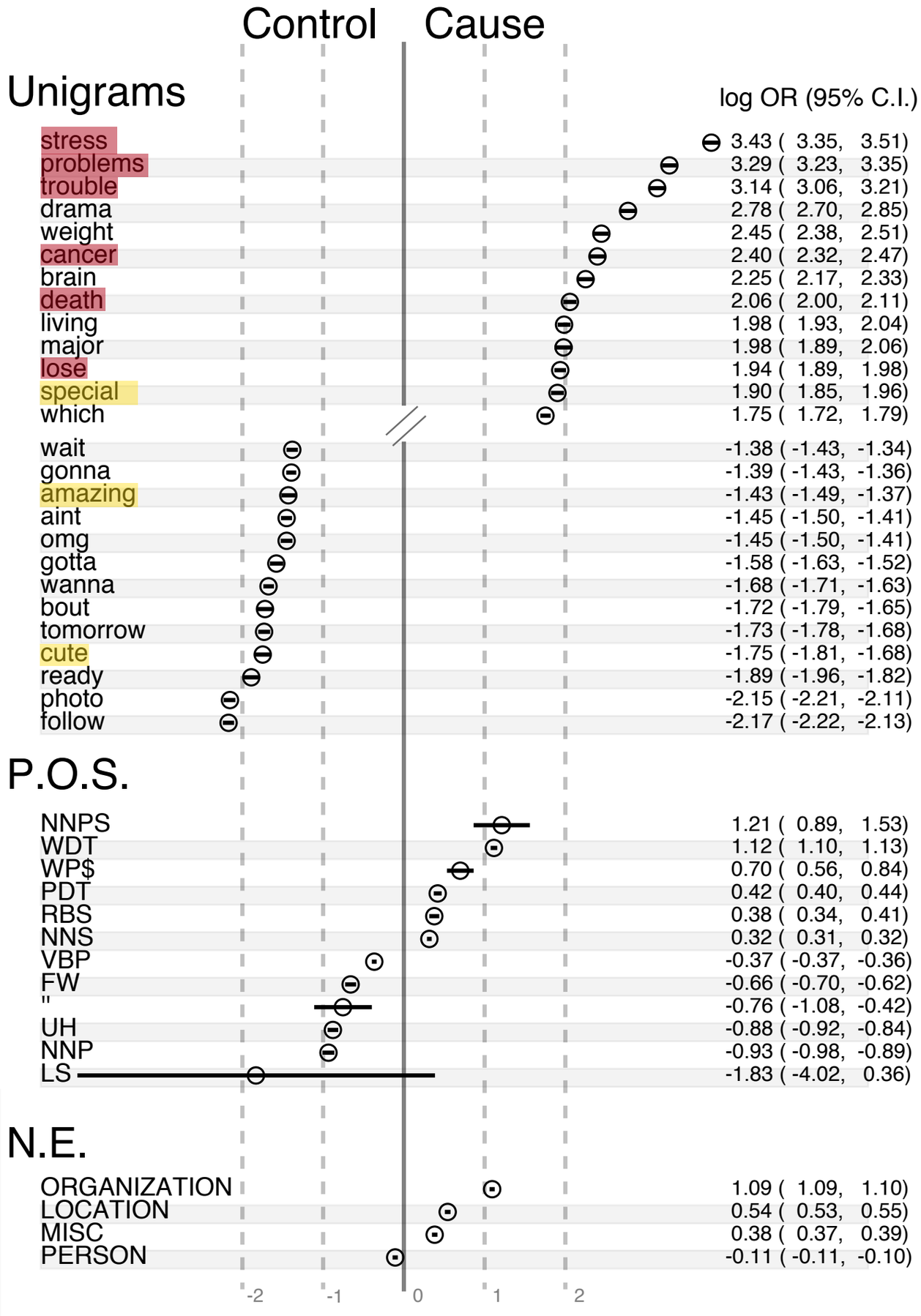
he$_{PRP}$ gon$_{MD}$ play$_{VB}$ until$_{IN}$ he$_{PRP}$ wins$_{VBZ}$ more$_{RBR}$ or$_{CC}$ until$_{IN}$ he$_{PRP}$ cant$_{VB}$ nomo$_{JJ}$ [...]

i$_{NN}$ think$_{VBP}$ the$_{DT}$ ps$_{NN}$ is$_{VBZ}$ a$_{DT}$ amazing$_{JJ}$ product$_{NN}$ its$_{PRP\$}$ worth$_{JJ}$ buying$_{NN}$

londons$_{O}$ appolo$_{O}$ theatre$_{O}$ collapses causing injuries eyewitnesses have described the chaos and panic

kevin$_{P}$ we have had tons of snow in bowermanville we could help the slopes at blue$_{L}$ mountain$_{L}$

Part of Speech (P.O.S)		**Named Entities (N.E.)**
WDT/ WP: Wh-determiner/pronoun		O:Organization
(N/NN)+(S/P): Noun plural/proper		L:Location
(P)+DT/CC:(pre)deter./conjunc.		P:Person
JJ+(S): adjective/superlative		M:Miscellaneous
MD/FW/LS:Modal/Foreign Word/List Item	VB+(GPNZ): Verb	RB+(RS): adverbs
(IN,PRP+$)/to: preposition (possess)/to	UH:Interjection	":quotes



i$_{NNS}$ actually$_{RB}$ hate$_{VBP}$ drama$_{NN}$ it$_{PRP}$ causes$_{VBZ}$ so$_{RB}$ much unnecessary$_{JJ}$ stress$_{NN}$

the$_{DT}$ bieber$_{NNP}$ family$_{NN}$ aka$_{VBZ}$ the$_{DT}$ cutest$_{JJS}$ family

this$_{DT}$ one$_{NN}$ problem$_{NN}$ has$_{VBZ}$ caused$_{VBN}$ so$_{RB}$ much$_{JJS}$ hurt$_{NN}$ and$_{CC}$ pain$_{NN}$ [...]

freezing$_{VBG}$ rain$_{NN}$ causes$_{NNS}$ thousands$_{NNS}$ to$_{TO}$ lose$_{VB}$ power$_{NN}$ across$_{IN}$ southern$_{JJ}$ ontario$_{NN}$

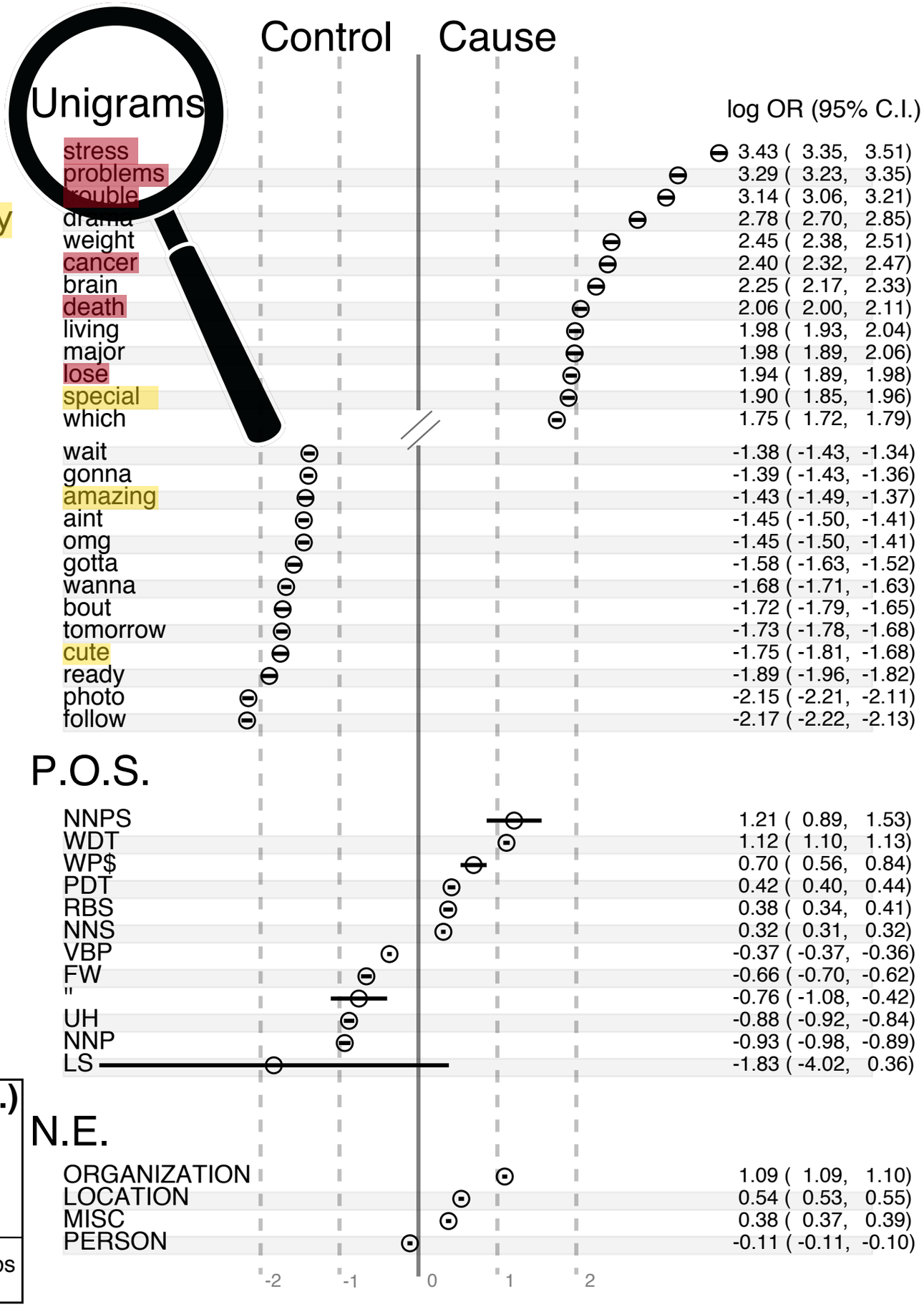
he$_{PRP}$ gon$_{MD}$ play$_{VB}$ until$_{IN}$ he$_{PRP}$ wins$_{VBZ}$ more$_{RBR}$ or$_{CC}$ until$_{IN}$ he$_{PRP}$ cant$_{VB}$ nomo$_{JJ}$ [...]

i$_{NN}$ think$_{VBP}$ the$_{DT}$ ps$_{NN}$ is$_{VBZ}$ a$_{DT}$ amazing$_{JJ}$ product$_{NN}$ its$_{PRP\$}$ worth$_{JJ}$ buying$_{NN}$

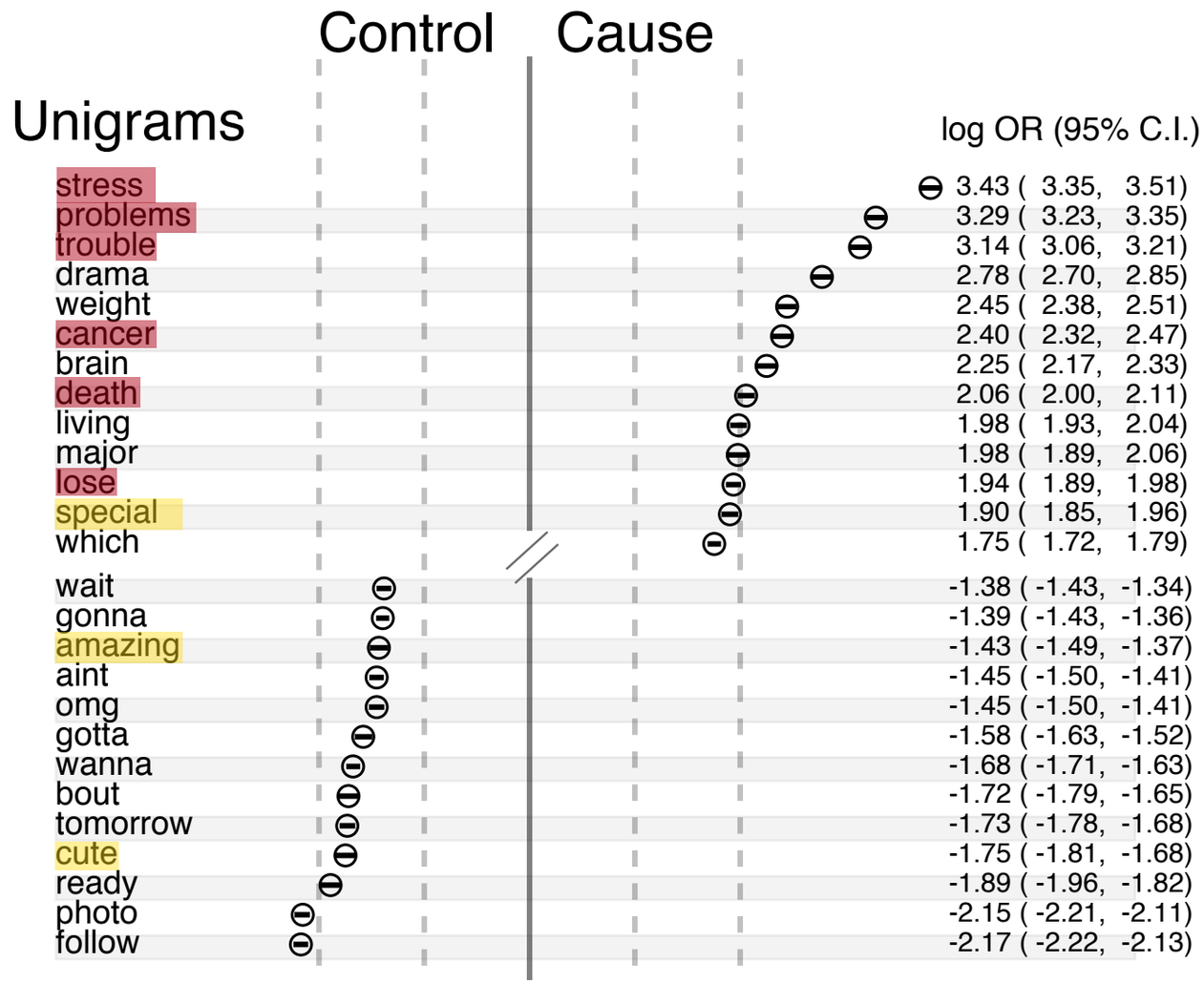
londons$_{O}$ appolo$_{O}$ theatre$_{O}$ collapses causing injuries eyewitnesses have described the chaos and panic

kevin$_{P}$ we have had tons of snow in bowermanville we could help the slopes at blue$_{L}$ mountain$_{L}$

Part of Speech (P.O.S)	**Named Entities (N.E.)**
WDT/ WP: Wh-determiner/pronoun	O:Organization
(N/NN)+(S/P): Noun plural/proper	L:Location
(P)+DT/CC:(pre)deter./conjunc.	P:Person
JJ+(S): adjective/superlative	M:Miscellaneous
MD/FW/LS:Modal/Foreign Word/List Item	VB+(GPNZ): Verb
(IN,PRP+$)/to: preposition (possess)/to	UH:Interjection
RB+(RS): adverbs	
":quotes	



Results
Unigrams



Part of Speech (P.O.S)

WDT/ WP: Wh-determiner/pronoun

(N/NN)+(S/P): Noun plural/proper

(P)+DT/CC:(pre)deter./conjunc.

JJ+(S): adjective/superlative

MD/FW/LS:Modal/Foreign Word/List Item

(IN,PRP+$)/to: preposition (possess)/to

VB+(GPNZ): Verb

UH:Interjection

RB+(RS): adverbs

":quotes

Named Entities (N.E.)

O:Organization

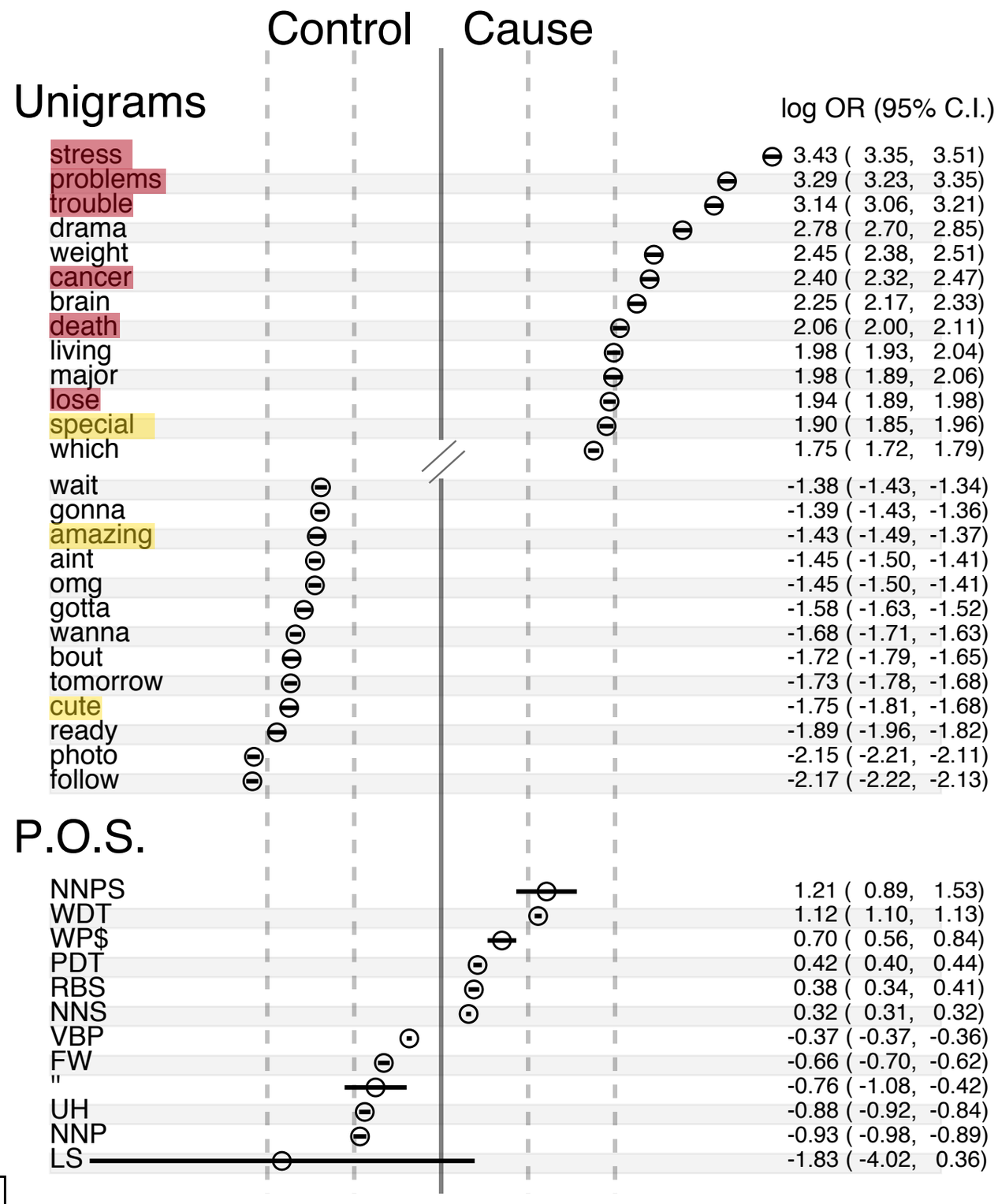
L:Location

P:Person

M:Miscellaneous

Results

Parts of Speech



Part of Speech (P.O.S)

WDT/ WP: Wh-determiner/pronoun

(N/NN)+(S/P): Noun plural/proper

(P)+DT/CC:(pre)deter./conjunc.

JJ+(S): adjective/superlative

MD/FW/LS:Modal/Foreign Word/List Item

(IN,PRP+$)/to: preposition (possess)/to

VB+(GPNZ): Verb

UH:Interjection

RB+(RS): adverbs

":quotes

Named Entities (N.E.)

O:Organization

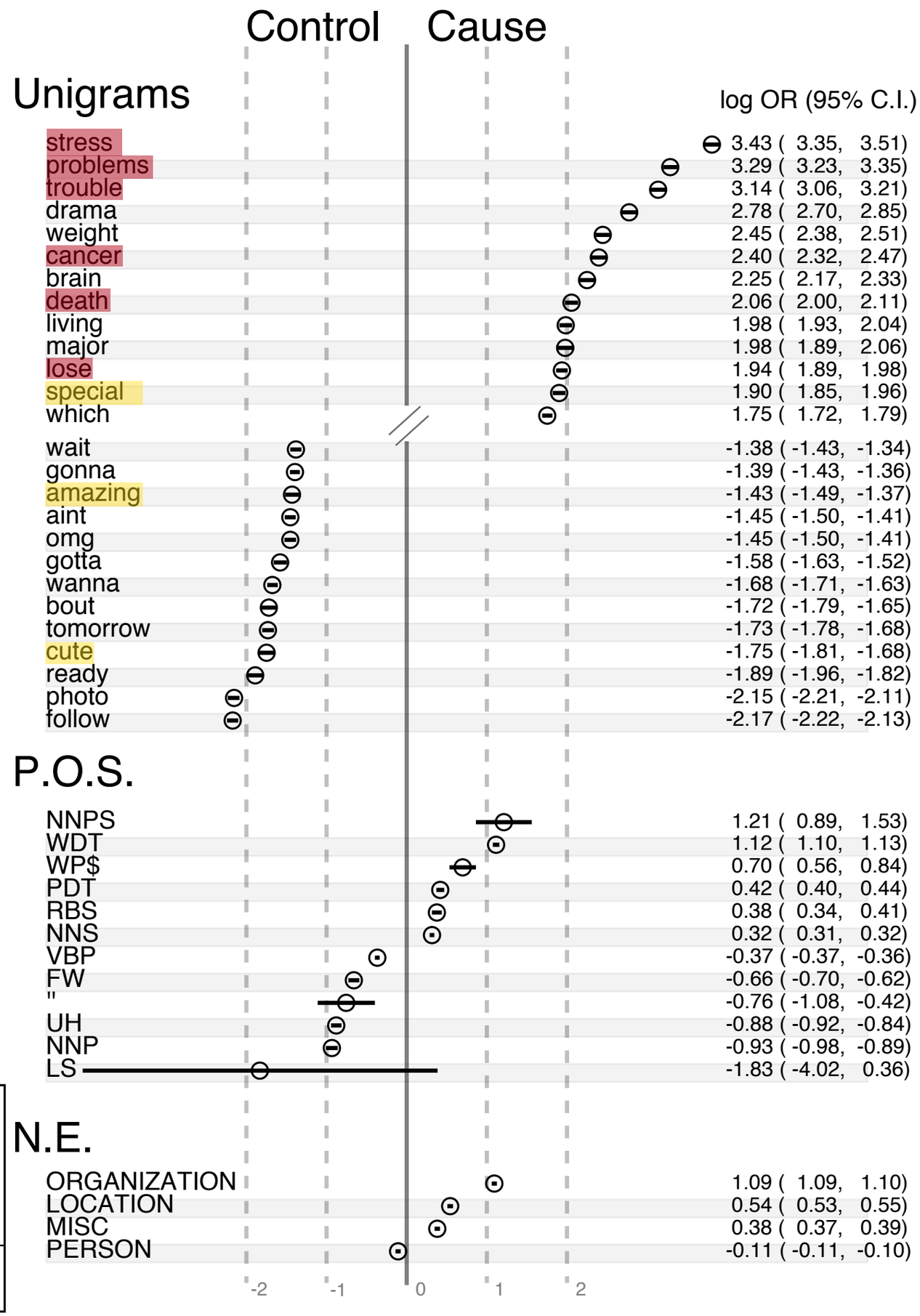
L:Location

P:Person

M:Miscellaneous

Results

Named Entities



Part of Speech (P.O.S)

WDT/ WP: Wh-determiner/pronoun

(N/NN)+(S/P): Noun plural/proper

(P)+DT/CC:(pre)deter./conjunc.

JJ+(S): adjective/superlative

MD/FW/LS:Modal/Foreign Word/List Item

(IN,PRP+$)/to: preposition (possess)/to

VB+(GPNZ): Verb

UH:Interjection

RB+(RS): adverbs

":quotes

Named Entities (N.E.)

O:Organization

L:Location

P:Person

M:Miscellaneous

i$_{NNS}$ actually$_{RB}$ hate$_{VBP}$ drama$_{NN}$ it$_{PRP}$ causes$_{VBZ}$ so$_{RB}$ much unnecessary$_{JJ}$ stress$_{NN}$

the$_{DT}$ bieber$_{NNP}$ family$_{NN}$ aka$_{VBZ}$ the$_{DT}$ cutest$_{JJS}$ family

this$_{DT}$ one$_{NN}$ problem$_{NN}$ has$_{VBZ}$ caused$_{VBN}$ so$_{RB}$ much$_{JJS}$ hurt$_{NN}$ and$_{CC}$ pain$_{NN}$ [...]

freezing$_{VBG}$ rain$_{NN}$ causes$_{NNS}$ thousands$_{NNS}$ to$_{TO}$ lose$_{VB}$ power$_{NN}$ across$_{IN}$ southern$_{JJ}$ ontario$_{NN}$

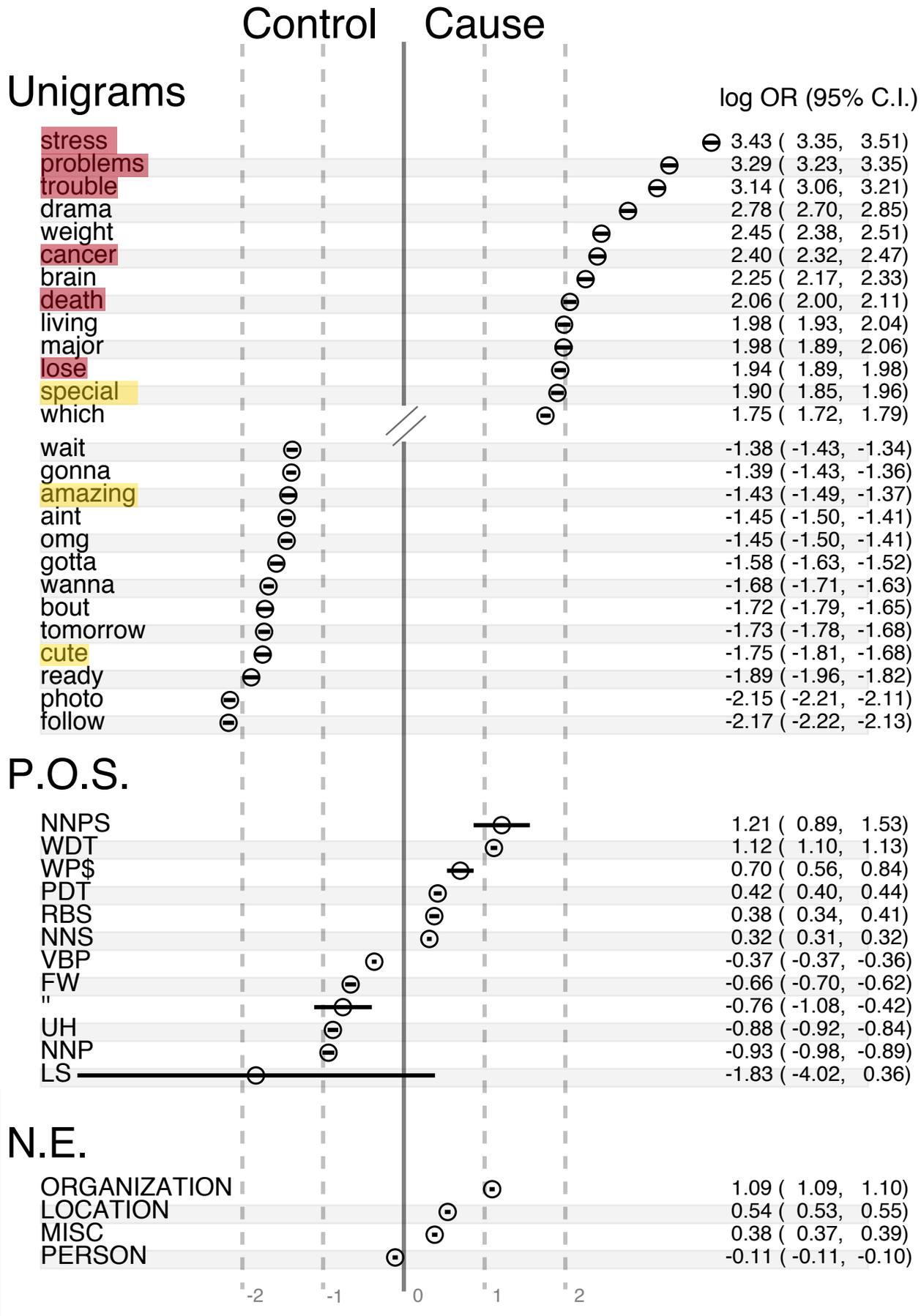
he$_{PRP}$ gon$_{MD}$ play$_{VB}$ until$_{IN}$ he$_{PRP}$ wins$_{VBZ}$ more$_{RBR}$ or$_{CC}$ until$_{IN}$ he$_{PRP}$ cant$_{VB}$ nomo$_{JJ}$ [...]

i$_{NN}$ think$_{VBP}$ the$_{DT}$ ps$_{NN}$ is$_{VBZ}$ a$_{DT}$ amazing$_{JJ}$ product$_{NN}$ its$_{PRP\$}$ worth$_{JJ}$ buying$_{NN}$

londons$_{O}$ appolo$_{O}$ theatre$_{O}$ collapses causing injuries eyewitnesses have described the chaos and panic

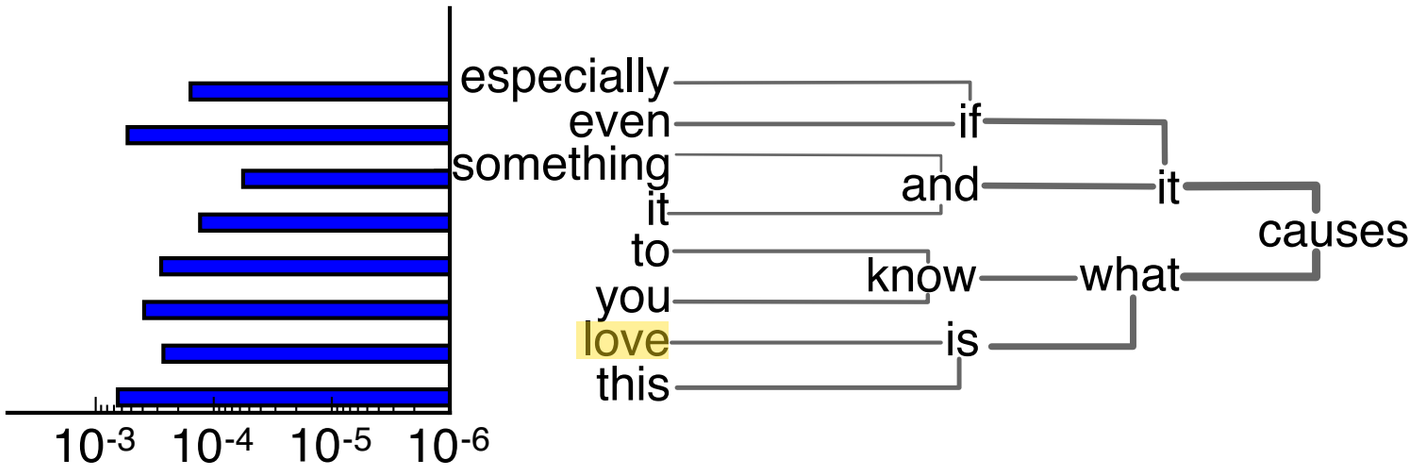
kevin$_{P}$ we have had tons of snow in bowermanville we could help the slopes at blue$_{L}$ mountain$_{L}$

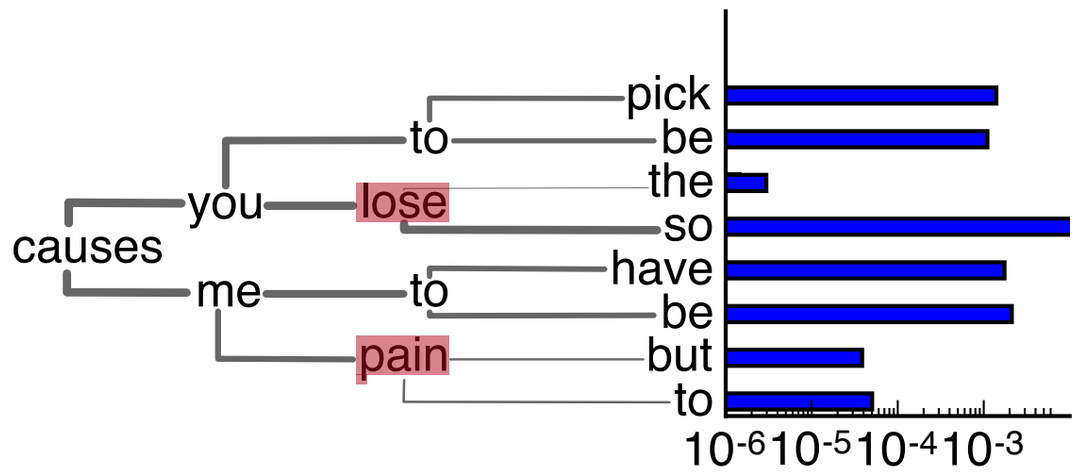
Part of Speech (P.O.S)		**Named Entities (N.E.)**
WDT/ WP: Wh-determiner/pronoun		O:Organization
(N/NN)+(S/P): Noun plural/proper		L:Location
(P)+DT/CC:(pre)deter./conjunc.		P:Person
JJ+(S): adjective/superlative		M:Miscellaneous
MD/FW/LS:Modal/Foreign Word/List Item	VB+(GPNZ): Verb	RB+(RS): adverbs
(IN,PRP+$)/to: preposition (possess)/to	UH:Interjection	":quotes

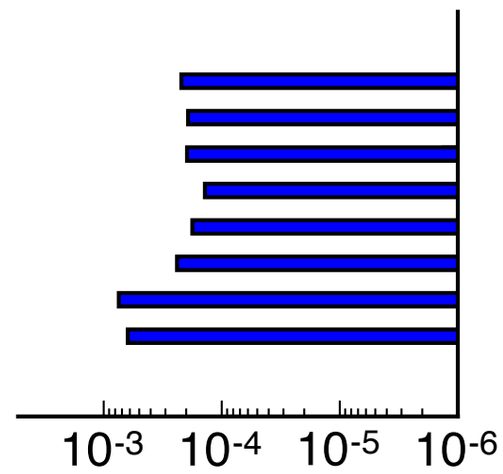


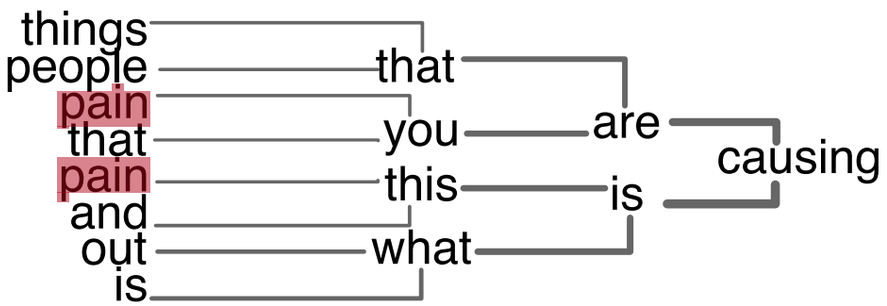
Results

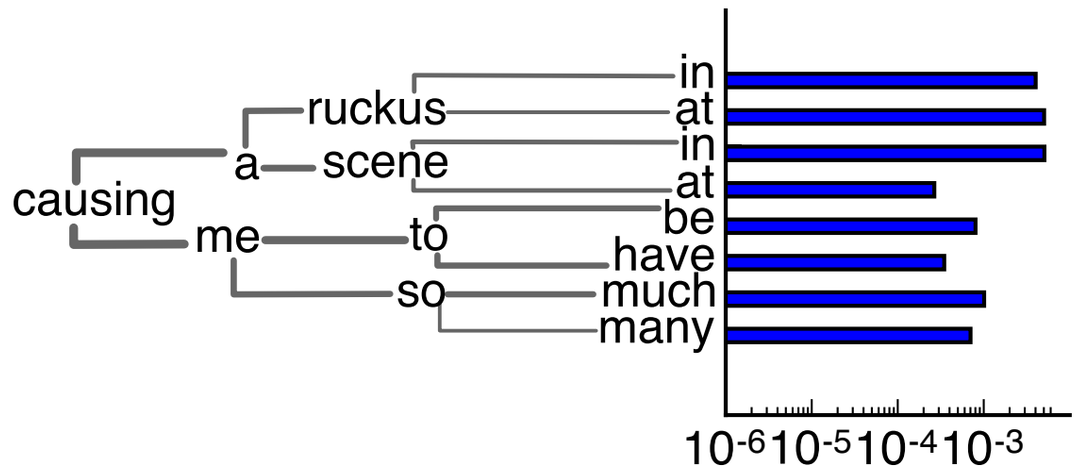
Cause-Trees

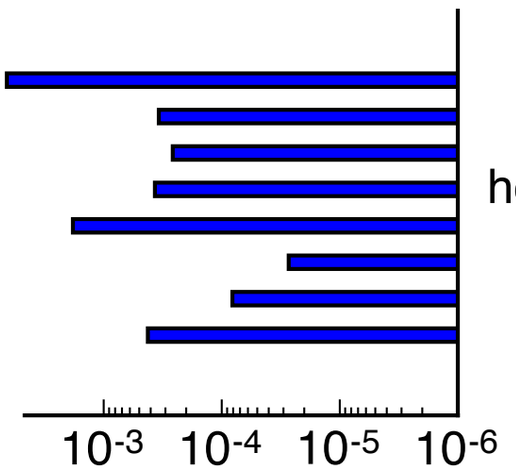


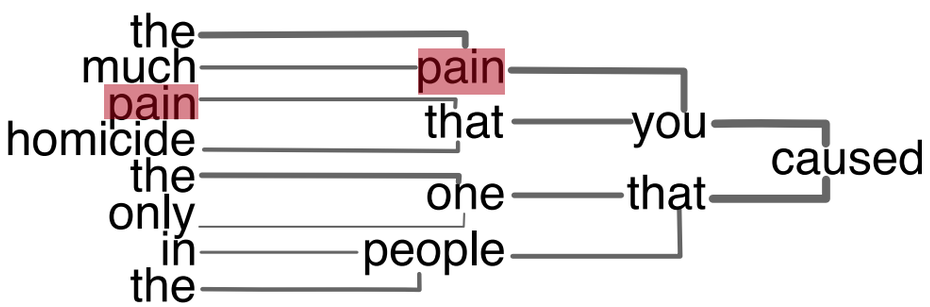


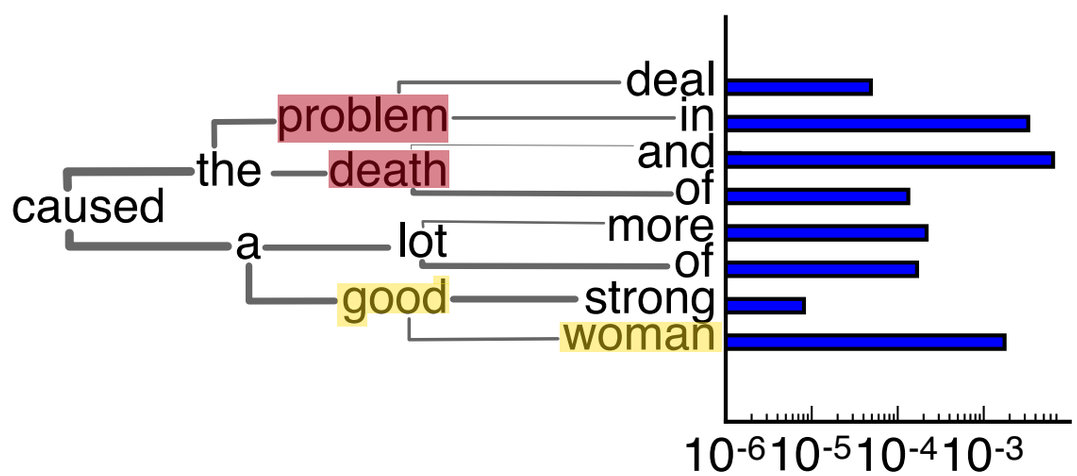






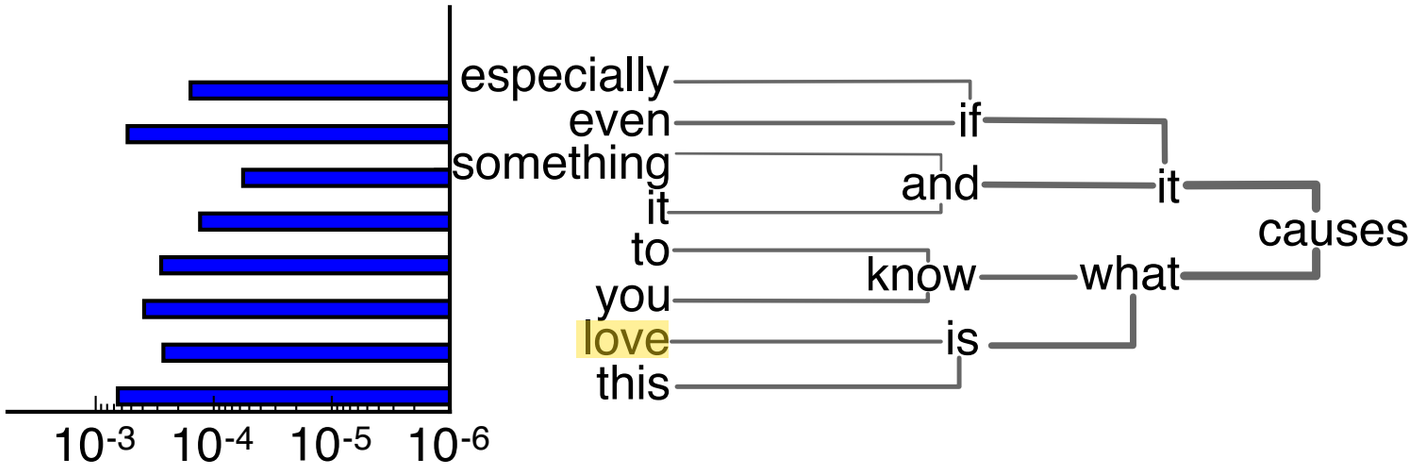


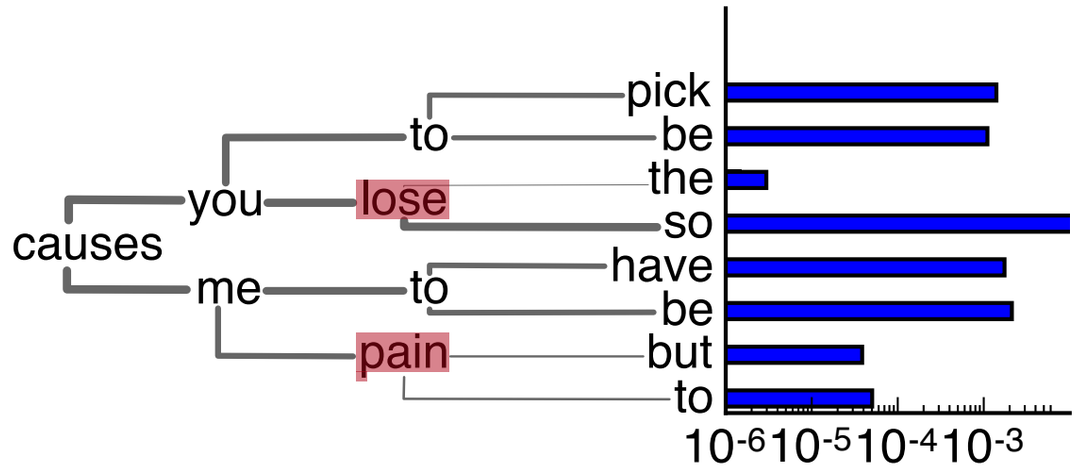




Results

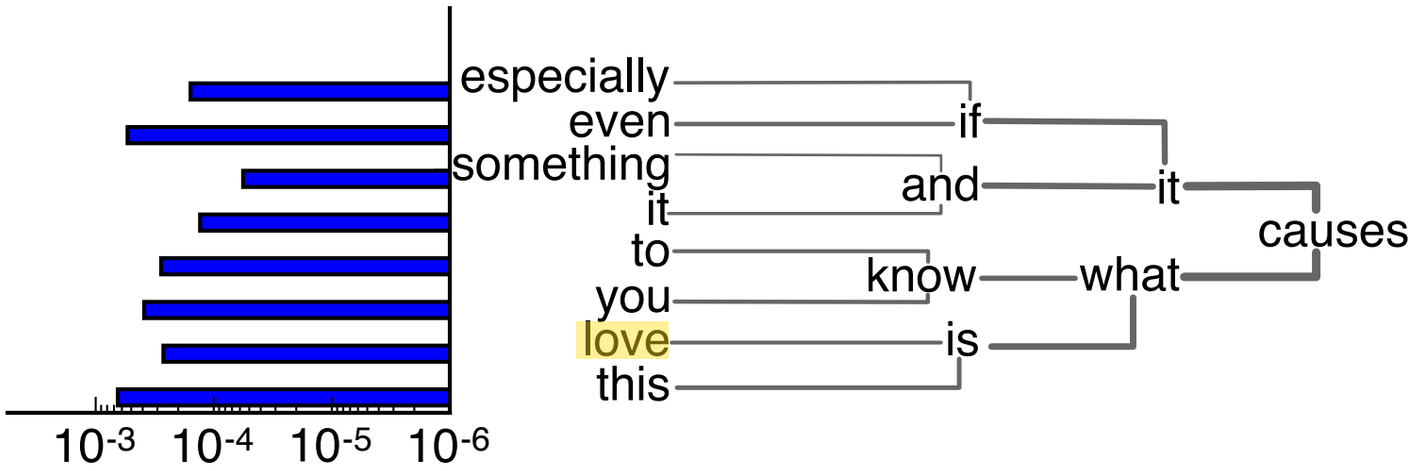
Cause-Trees

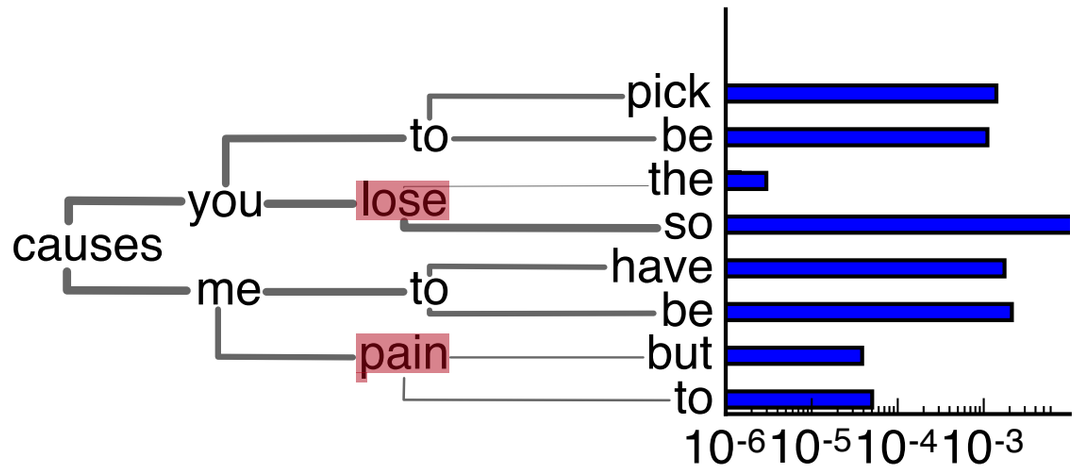


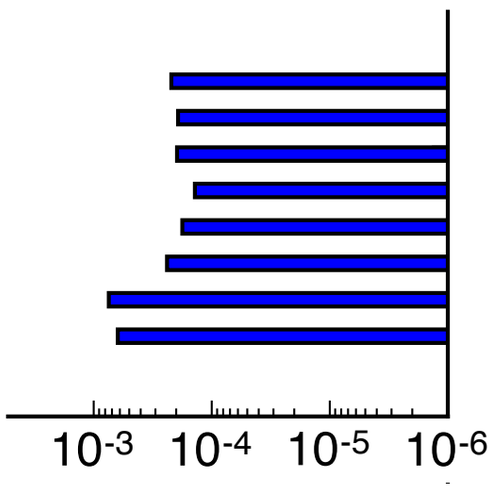


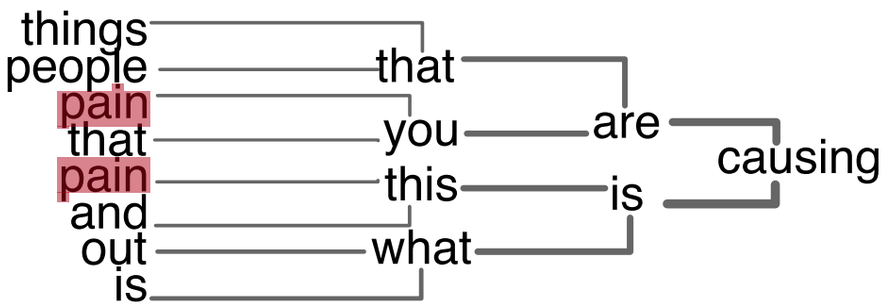
Results

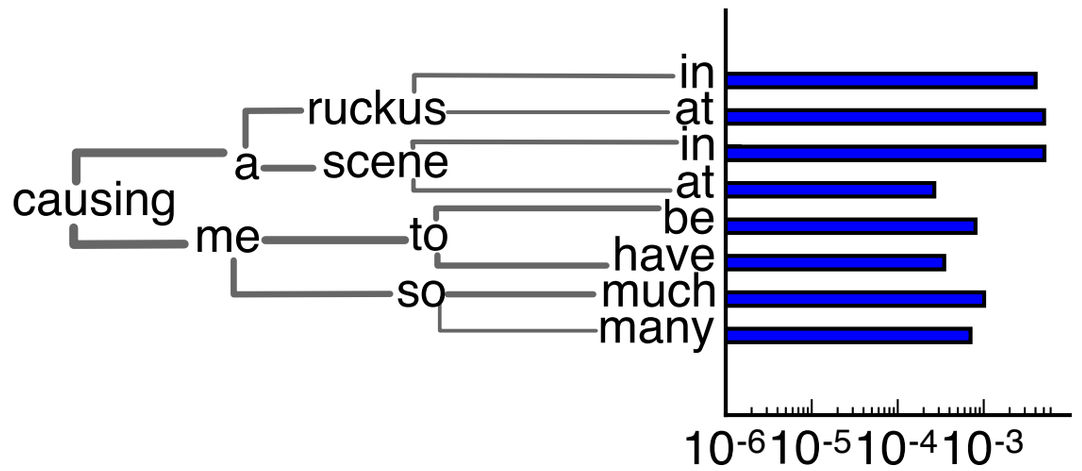
Cause-Trees





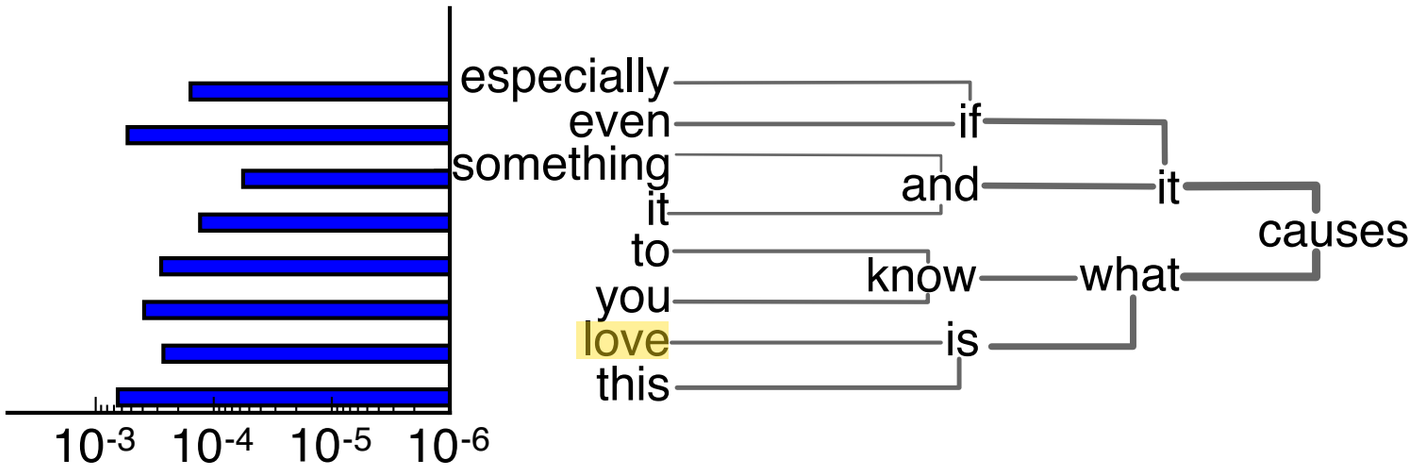


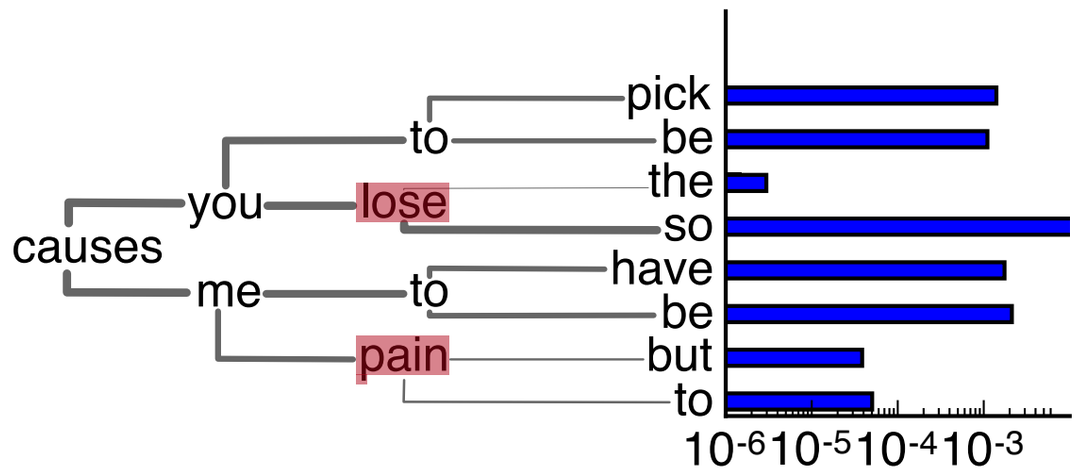


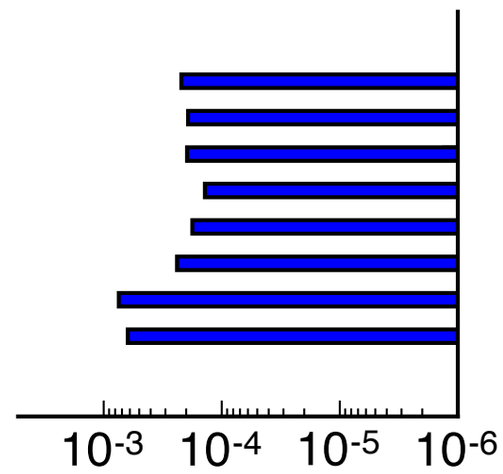


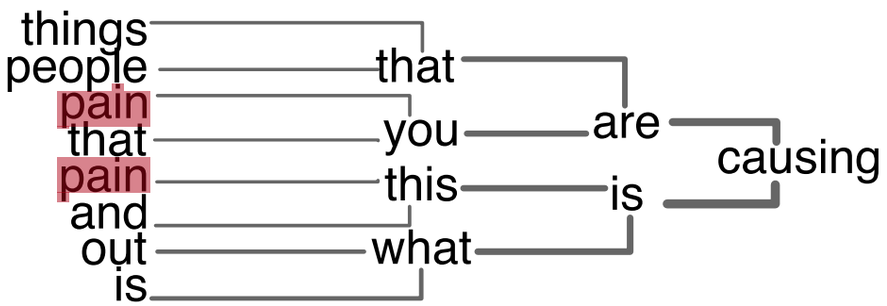
Results

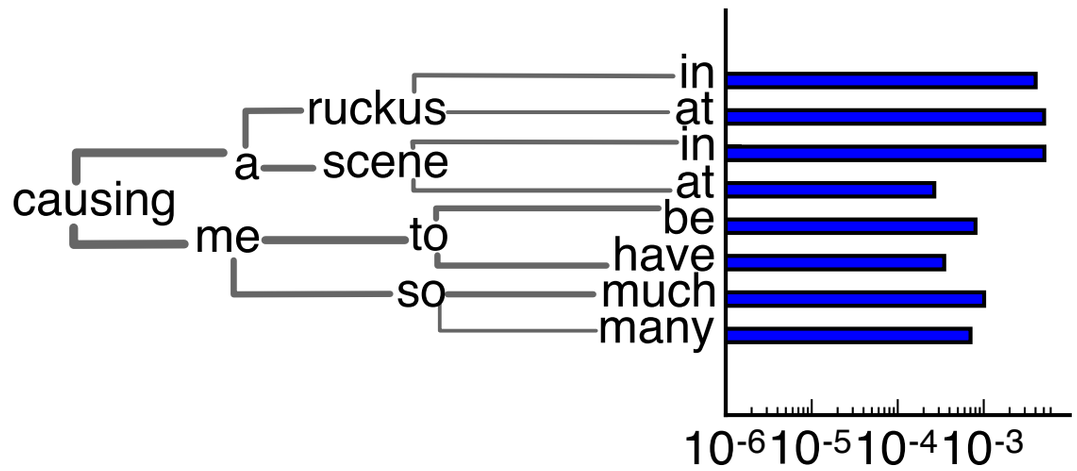
Cause-Trees

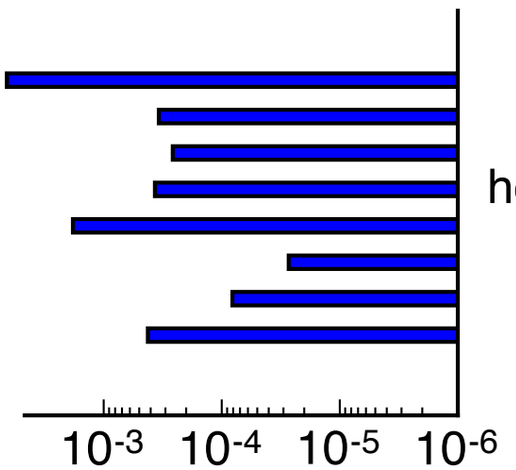


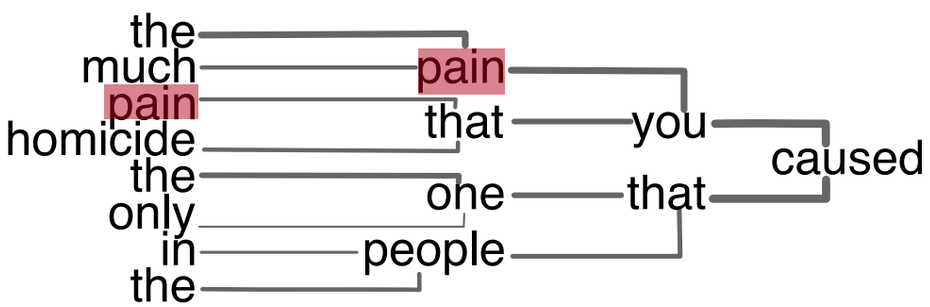


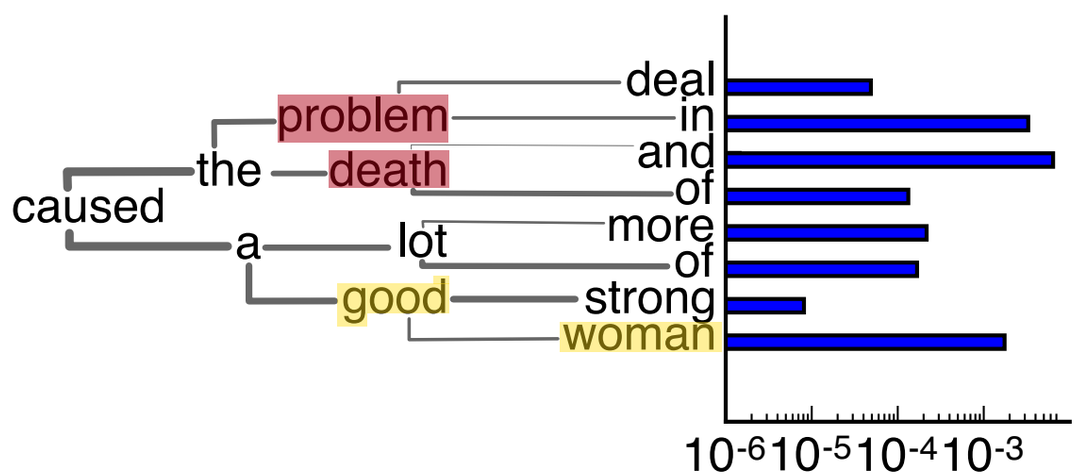












Results

Topic Modeling

"News"	"Accident"	"Problems"	"Medical"	"Crisis"	"Sorry"	"Stress"	"Body"	"Drama"	"Injuries"
damage	traffic	dont	cancer	their	any	more	stress	like	his
fire	delays	people	break	our	never	than	lose	she	him
power	crash	they	some	from	been	being	weight	her	out
via	car	problems	men	how	sorry	over	stuff	lol	back
new	accident	why	can	about	there	person	living	out	her
news	death	about	disease	social	know	sleep	quickly	trouble	when
from	between	when	from	crisis	will	which	special	good	head
says	after	know	most	via	ive	people	proof	now	into
after	year	them	our	great	they	one	diets	sh*t	off
video	down	like	others	money	out	stress	excercise	life	well
global	there	drama	loss	many	problems	someone	f*ck	twitter	which
rain	man	who	heart	issues	can	makes	who	got	from
warming	due	one	health	should	now	think	giving	scene	down
water	from	youre	food	war	trouble	when	love	get	game
explosion	snow	get	symptoms	problems	see	most	god	too	fall
outage	road	stop	hair	government	one	thinking	people	girl	face
storm	old	think	women	true	how	brain	will	haha	then
change	over	how	blood	new	would	depression	our	needs	get
house	problems	sh*t	how	world	could	anxiety	those	see	injuries
may	chaos	want	skin	they	were	lack	one	walk	had
flooding	morning	cant	records	media	had	night	around	drama	sports
gas	two	because	adversity	other	ever	without	life	hes	stick
air	driving	too	high	obama	whats	love	his	woman	over
say	major	hate	helium	financial	again	mental	thats	some	while
stir	today	need	which	change	did	them	work	last	eyes
heavy	disruption	only	eating	violence	time	mind	out	strong	only
weather	train	really	may	will	think	fact	good	shes	hit
collapse	accidents	many	body	also	well	insomnia	sex	always	famous
climate	almost	even	smoking	shutdown	something	hand	come	him	hockey

TABLE I

Results

Topic Modeling - News

"News"	"Accident"	"Problems"
damage	traffic	dont
fire	delays	people
power	crash	they
via	car	problems
new	accident	why
news	death	about
from	between	when
says	after	know
after	year	them
video	down	like
global	there	drama
rain	man	who
warming	due	one
water	from	youre
explosion	snow	get
outage	road	stop
storm	old	think
change	over	how
house	problems	sh*t
may	chaos	want
flooding	morning	cant
gas	two	because
air	driving	too
say	major	hate
stir	today	need
heavy	disruption	only
weather	train	really
collapse	accidents	many
climate	almost	even

TABLE I

Results

Topic Modeling - Emergency

"News"	"Accident"	"Problems"	"Medical"	"Crisis"
damage	traffic	dont	cancer	their
fire	delays	people	break	our
power	crash	they	some	from
via	car	problems	men	how
new	accident	why	can	about
news	death	about	disease	social
from	between	when	from	crisis
says	after	know	most	via
after	year	them	our	great
video	down	like	others	money
global	there	drama	loss	many
rain	man	who	heart	issues
warming	due	one	health	should
water	from	youre	food	war
explosion	snow	get	symptoms	problems
outage	road	stop	hair	government
storm	old	think	women	true
change	over	how	blood	new
house	problems	sh*t	how	world
may	chaos	want	skin	they
flooding	morning	cant	records	media
gas	two	because	adversity	other
air	driving	too	high	obama
say	major	hate	helium	financial
stir	today	need	which	change
heavy	disruption	only	eating	violence
weather	train	really	may	will
collapse	accidents	many	body	also
climate	almost	even	smoking	shutdown

TABLE I

Results

Topic Modeling - Social

"News"	"Accident"	"Problems"	"Medical"	"Crisis"	"Sorry"	"Stress"	"Body"	"Drama"	"Injuries"
damage	traffic	dont	cancer	their	any	more	stress	like	his
fire	delays	people	break	our	never	than	lose	she	him
power	crash	they	some	from	been	being	weight	her	out
via	car	problems	men	how	sorry	over	stuff	lol	back
new	accident	why	can	about	there	person	living	out	her
news	death	about	disease	social	know	sleep	quickly	trouble	when
from	between	when	from	crisis	will	which	special	good	head
says	after	know	most	via	ive	people	proof	now	into
after	year	them	our	great	they	one	diets	sh*t	off
video	down	like	others	money	out	stress	excercise	life	well
global	there	drama	loss	many	problems	someone	f*ck	twitter	which
rain	man	who	heart	issues	can	makes	who	got	from
warming	due	one	health	should	now	think	giving	scene	down
water	from	youre	food	war	trouble	when	love	get	game
explosion	snow	get	symptoms	problems	see	most	god	too	fall
outage	road	stop	hair	government	one	thinking	people	girl	face
storm	old	think	women	true	how	brain	will	haha	then
change	over	how	blood	new	would	depression	our	needs	get
house	problems	sh*t	how	world	could	anxiety	those	see	injuries
may	chaos	want	skin	they	were	lack	one	walk	had
flooding	morning	cant	records	media	had	night	around	drama	sports
gas	two	because	adversity	other	ever	without	life	hes	stick
air	driving	too	high	obama	whats	love	his	woman	over
say	major	hate	helium	financial	again	mental	thats	some	while
stir	today	need	which	change	did	them	work	last	eyes
heavy	disruption	only	eating	violence	time	mind	out	strong	only
weather	train	really	may	will	think	fact	good	shes	hit
collapse	accidents	many	body	also	well	insomnia	sex	always	famous
climate	almost	even	smoking	shutdown	something	hand	come	him	hockey

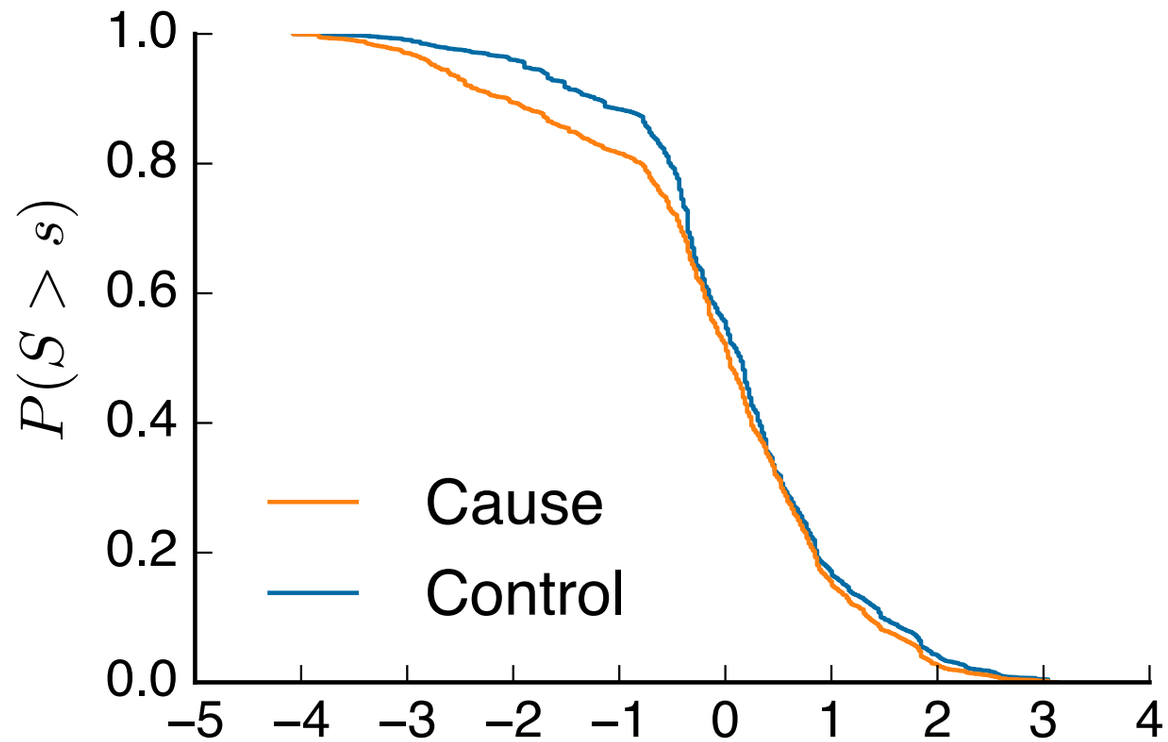
TABLE I

A

Filtered

	Cause	Control
mean	-0.105	0.111
SE	4.562×10^{-4}	4.180×10^{-4}

B

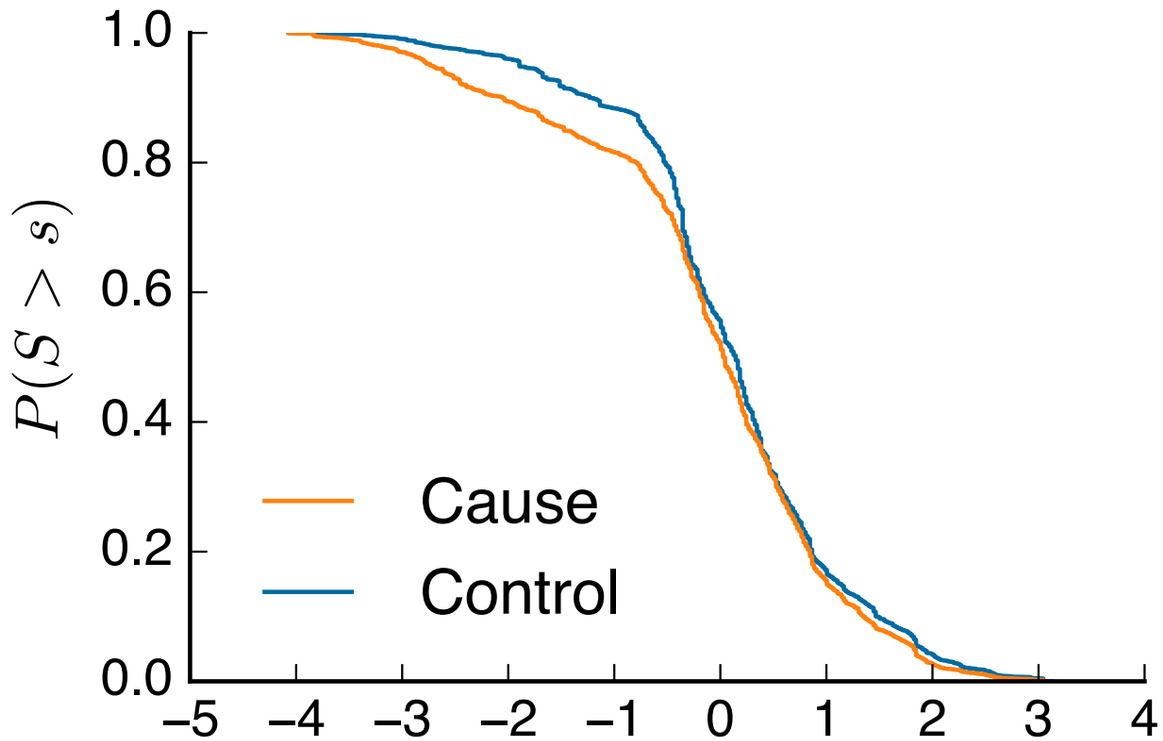


A

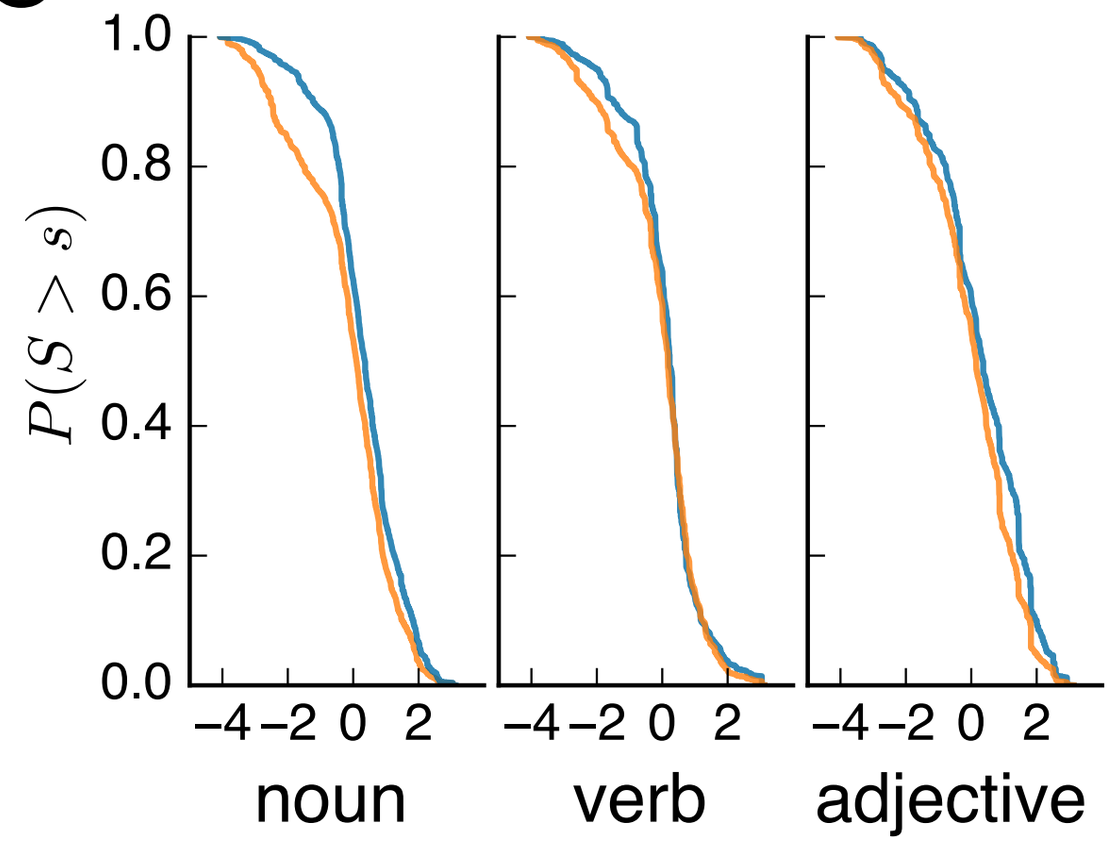
Filtered

	Cause	Control
mean	-0.105	0.111
SE	4.562×10^{-4}	4.180×10^{-4}

B



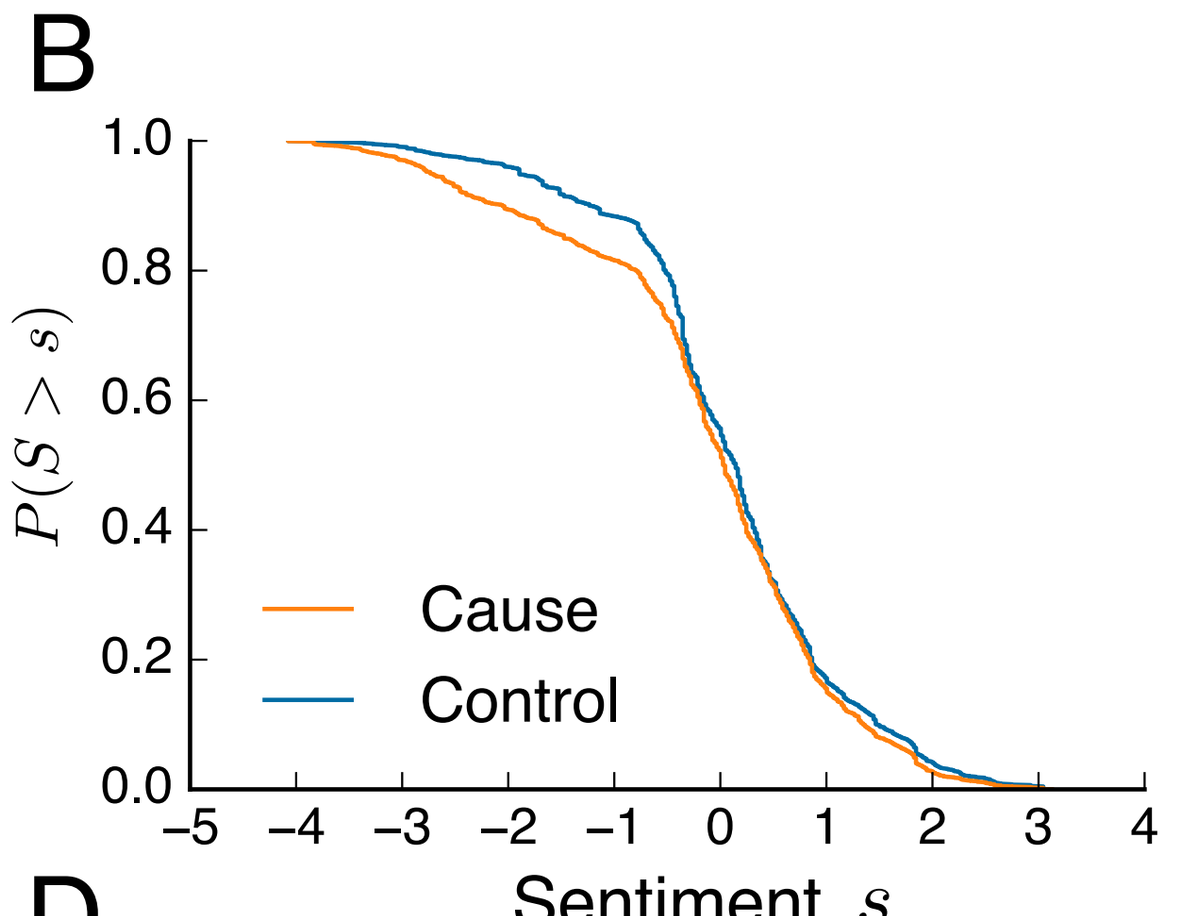
C

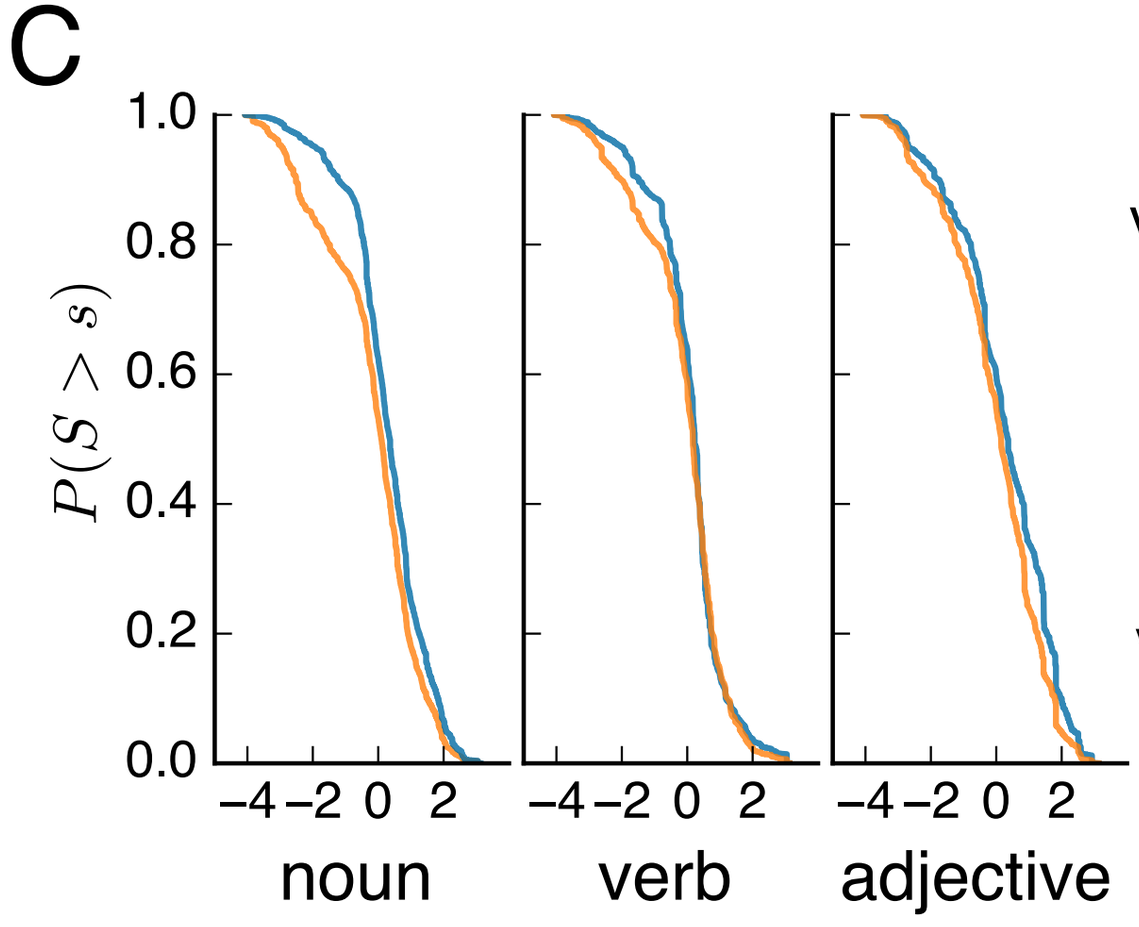


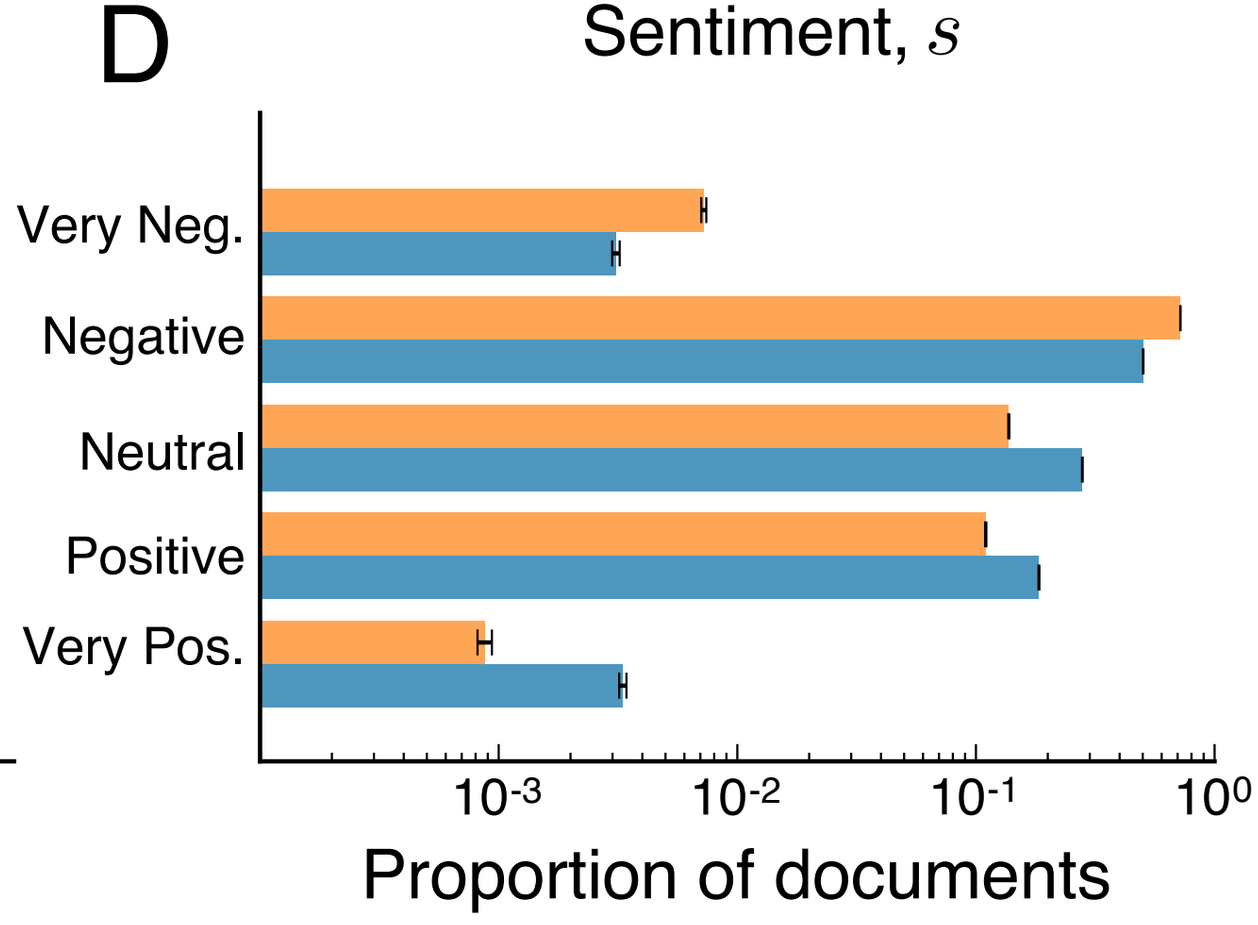
A

Filtered

	Cause	Control
mean	-0.105	0.111
SE	4.562×10^{-4}	4.180×10^{-4}







Thank you

&

Questions

Collaborators

James Bagrow

Josh Bongard

Chris Danforth

Paul Hines

Peter Dodds