

“Average number of words
in the written vocabulary
of a 6- to 14-year-old
American child in 1945: **25,000**”

“Average number of words
in the written vocabulary
of a 6- to 14-year-old
American child in 1945:

25,000

“Average number today:

10,000”

“Average number of words
in the written vocabulary
of a 6- to 14-year-old
American child in 1945: **25,000**
“Average number today: **10,000**”

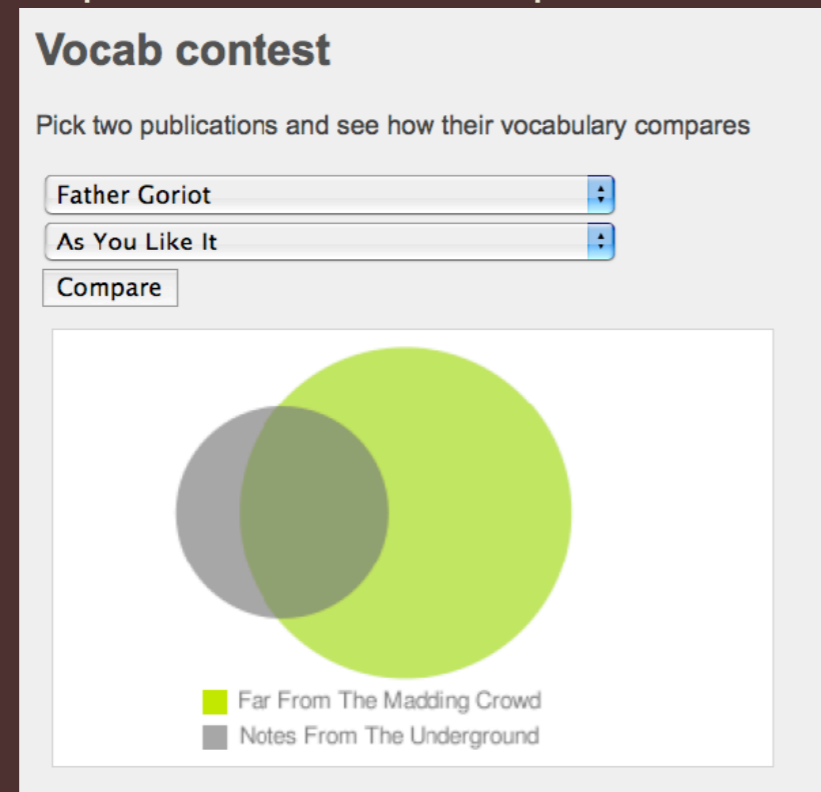
So reported *Harper's* in August 1990.¹ The validity of this comparison has been called into question as being one of not apples to apples, but rutabegas to rutabegas,² yet our curiosity about the matter remains: ***How has our vocabulary changed over time?***

1 Index, *Harper's*, August 1990, <http://is.gd/iWaf>

2 “Does the average American student have less vocabulary today than in days gone by?” Cecil Adams, *The Straight Dope*, September 7, 2007, <http://is.gd/iWbZ>

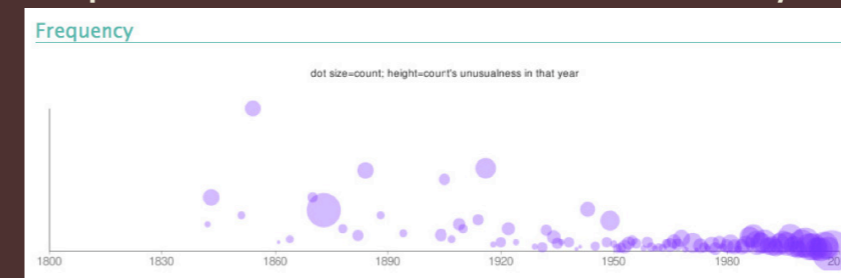
My project is to **compare the vocabularies** of a collection of English-language novels published between the late nineteenth century and the present day, in order to measure the overlap between the those vocabularies, as well as to identify specific words that are exclusive to one group or another.

Sample visualization from Book Scraper



<http://labs.timesonline.co.uk/bookscraper/>

Sample visualization from Wordnik of the word "vanity"



<http://www.wordnik.com/>

Allen, James
Lane
*The Reign of
Law*

Bachelor,
Irving
D' Ri and I

Bachelor,
Irving
Eben Holden

Bova, Ben
*The Sam Gunn
Omnibus*

Brust, Steven
*My Own Kind
of Freedom*

Bryan,
Kathleen
*The Serpent
and the Rose*

Chestnutt,
Charles W
*The House
Behind the
Cedars*

Churchill,
Winston
Richard Carvel

de Lint,
Charles
Spiritwalk

Doctorow,
Cory
*Down and Out
in the Magic
Kingdom*

Doctorow,
Cory
*Eastern Stan-
dard Tribe*

Docto
Cory
*Some
Come
Town
one L
Town*

**Drag books here
to compare them**

**Drag books here
to compare them**

Bova, Ben <i>Sam Gunn Omnibus, The</i>	Brust, Steven <i>My Own Kind of Freedom</i>	Bryan, Kathleen <i>The Serpent and the Rose</i>	de Lint, Charles <i>Spiritwalk</i>	Dorfman, Ariel <i>Mascara</i>	Flint, Eric <i>1632</i>	Flint, Eric <i>1633</i>	Flint, Eric <i>1812: The Rivers of War</i>	Gifford, Barry <i>Wyoming</i>	Guon, Ellen <i>Bedlam Boyz</i>	Horne, Marc <i>Tokyo Zero</i>	Stros Char Acce
---	--	--	---------------------------------------	----------------------------------	----------------------------	----------------------------	---	----------------------------------	-----------------------------------	----------------------------------	-----------------------

Doctorow, Cory <i>Down and Out in the Magic Kingdom</i>	Doctorow, Cory <i>Eastern Standard Tribe</i>	Doctorow, Cory <i>Someone Comes to Town Someone Leaves Town</i>	Duberman, Martin <i>Haymarket</i>
Franco, Jorge <i>Rosario Tijeras</i>	Lackey, Mercedes <i>Beyond World's End</i>	Link, Kelly <i>Stranger Things Happen</i>	Mamatas, Nick <i>Move Under Ground</i>
John Ringo <i>A Hymn Before Battle</i>	Watts, Peter <i>Behemoth</i>		

Total words: 885,656 • Unique: 37,542 • Only in this group: 13,780



Allen, James Lane <i>The Reign of Law</i>	Bacheller, Irving <i>D'ri and I</i>	Bacheller, Irving <i>Eben Holden</i>	Chestnutt, Charles W <i>The House Behind the Cedars</i>
Churchill, Winston <i>Richard Carvel</i>	Dreiser, Theodore <i>Sister Carrie</i>	Ford, Paul Leicester <i>Janice Meredith</i>	Goss, Charles Frederic <i>The Redemption of David Corson</i>
Grant, Robert <i>Unleavened Bread</i>	Johnston, Mary <i>Audrey</i>	Johnston, Mary <i>To Have and to Hold</i>	Major, Charles <i>Dorothy Vernon of Haddon Hall</i>
Major, Charles <i>When Knight-</i>	Norris, Frank <i>The Octopus</i>	Tarkington, Booth <i>The Two Van-</i>	Thompson, Maurice

Total words: 885,656 • Unique: 37,542 • Only in this group: 16,719



Bova, Ben
Sam Gunn Omnibus, The

Brust, Steven
My Own Kind of Freedom

Bryan, Kathleen
The Serpent and the Rose

de Lint, Charles
Spiritwalk

Doctorow, Cory
Eastern Standard Tribe

Dorfman, Ariel
Mascara

Duberman, Martin
Haymarket

Flint, Eric
1632

Flint, Eric
1633

Flint, Eric
1812: The Rivers of War

Franco, Jorge
Rosario Tijeras

Gifford, Wyo

Doctorow, Cory
Down and Out in the Magic Kingdom

absentmindedly ac accreting accrued achingly acquisitive addled
adrenaline adult adulthood adults adventureland afro agenda ahold
airhead airline airlock airtyped aka alibi alibis allocate aluminum ambient
ambivalent ambling animatronic animatronics annotations anonymity
anthro anymore anytime app apparati arcane archived arguably
armchairs armored asap assed assessed asshole asteroid atcha audio
audiotrack authorization autohypnotic autopilot autosolder avatars awards
backdrop backside backstage backtraced backup backups badmouthing
balancer barrelhouse basically basketball bathtub bazaar beater beatles
beeline beers behaviors beijing belched bellycrawl bemused beneficently
beta betcha bicep biceps bifocals bile billions biometrically bioscience
bitching bitchun bitchunry blackjack blaster bleary blindspot bloodstream
boggled bolstering boo booga bopped bossing bot bothering bottleneck
bounced bowsprits boxers brainwashing brandished breakdown
breathtaking breezeway briefed broadband brrrr btw bubbly buddies
buddy bugged bulges bullshit bumped bummed bun bungalow bunking
bunnies burgeoning bursitis busses buttocks bzzz cadge cadre cagey
callused calluses camaraderie cameras campground campsite canaveral
canopic captions caribbean castmember castmembers castmemberspeak
castmembery catchy catheterized cavemen ceded cheekbones cheeriest
chem chemistry chend cheyenne chills chirpy chitchat chortled chrissakes
chub chum chutzpah cicadas cig civvies clarity claustrophobic cleavage
clipboard cliques clone clones closures clunk coaster cobbled cochlea
cocking cocktail codes coding coed cogitated coitus collaborating
collaborative collarbones collectible collective commitment commitments
commune compulsive compulsively computer computers conceptual
concordia concurred conditioning condo cons consciousnesses consensus
conservatives contacted contacts contrariwise conundra convocation

Total words: 47,344 • Unique: 6,915 • Only in this book: 1,350



Allen, James Lane
The Reign of Law

Bacheller, Irving
D'ri and I

Bacheller, Irving
Eben Holden

Chestnutt, Charles W
The House Behind the Cedars

Churchill, Winston
Richard Carvel

Dreiser, Theodore
Sister Carrie

Ford, Paul Leicester
Janice Meredith

Goss, Charles Frederic
The Redemption of David Corson

Grant, Robert
Unleavened Bread

Johnston, Mary
Audrey

Johnston, Mary
To Have and to Hold

Major, Charles
Dorothy Vernon of Haddon Hall

Major, Charles
When Knight-

Norris, Frank
The Octopus

Tarkington, Booth
The Two Van-

Thompson, Maurice

Total words: 885,656 • Unique: 37,542 • Only in this group: 30,749



Allen, James
Lane
*The Reign of
Law*

Bachelor,
Irving
D'ri and I

Bachelor,
Irving
Eben Holden

Bova, Ben
*Sam Gunn
Omnibus, The*

Brust, Steven
*My Own Kind
of Freedom*

Bryan,
Kathleen
*The Serpent
and the Rose*

Chestnutt,
Charles W
*The House
Behind the
Cedars*

Churchill,
Winston
Richard Carvel

de Lint,
Charles
Spiritwalk

Doctorow,
Cory
*Eastern Stan-
dard Tribe*

Dorfman, Ariel
Mascara

Dube
Mart
Hayr

Doctorow, Cory *Down and Out in the Magic Kingdom*

abandoning abbreviated abdicate abe absentmindedly ac accelerated accents access accomplice accounting accreting accrue accrued accumulate accumulating accurately accusations achingly acquisitive additional addled admixture adolescence adrenaline adrift adult adulthood adults advanced adventureland adventurer adventurers adversity advocating affirmations affording afoot afresh africa afro agenda ages aggregate ahold ai ailing airhead airline airlock airtyped aka alarmed albeit alibi alibis alienating allies allocate ally almighty alongside aloud alternate alternated alternately aluminum amazed amazing ambient ambivalent ambling anatomy angled animatronic animatronics ankles annotations anonymity answers anthro anymore anyone anytime apartment aping aplenty apologized app apparati apparatus appearances apple appreciatively apprehensively aquatic arcade arcane archived areas arguably argumentative arisen arithmetic armchairs armies armor armored armpit articulated asap ascent asia aspects aspiring assassin assassination assassinations assed asses assessed asshole assistance assorted asteroid atcha atom attempts attire attributed au audacious audio audiotrack auditorium authorization autohypnotic automatic automatically autopilot autosolder availability avatars averted averting awards axes backdrop backside backstage backtraced backup backups backwards badmouthing bag bait baited bake baked baking balancer balled balm banana banging banished banked barbecue bargain barge barges baring bark barked barrel barrelhouse barrels basic basically basketball bastard bastards batch bathtub battering bazaar beach beaded beam beamed bears beastly beater beatles beeline beer beers behavior behaviors beijing belched beliefs belligerent belly bellycrawl beloved bemused ben bend beneficently benighted beta betcha betray betrayed betsy bewildering bicep biceps bicycles bifocals bikes bile billions

Total words: **47,344** • Unique: **6,915** • Only in this book: **3,196**



Dreiser, Theodore *Sister Carrie*

aback abandonment abandons abash abashed abashing abated abdul abide abilities ablaze ably aboard abode abodes abolished abroad absence absent absorbed absorbing absorbs absorption abstractedly abundance abuse abused accept acceptably accepted accepting accessories accident accidentally acclimatization accommodations accomplishments accordingly accosted account accoutrements accretion accusation accusing achieved achievement aching acknowledgment acquaintance acquaintances acquaintanceship acquainted acquiesce acquiesced acquiescence acquitted actively actor actors actress actresses actual actualities actuated actuating acuteness adam adams adapted addition addressed addresses addressing adjective adjoining adjuncts adjust adjustable adjustment administer admirable admirably admiration admire admirers admiringly admissible admitting ado adopt adopts adored adorer adornment adroitness ads adulation advance advancement advances advantage advantageous advantages adventures adventuresome advertise advertised advertisement advertisements advisability advise advised advising advocated afar affability affable affably affair affairs affect affectation affectations affected affecting affection affectional affirmation affixed afflatus afflicted affluence affluent affray afternoons afterthought afterward agates aged agencies agent aggravated aggravation aggrieved agility agitated agitation aglow agonising agony agreeable agreeably agreement aid aided aiding ailed ailment aim aiming aimless aims airily airiness airing airs airy aisle akimbo alacrity aladdin alarming alas albert alchemy alcove alderman aldermanic ale alight alighted align aligned alimony alive alleged allied alligator allotted allowances allows allure allured allurements allures alluring alms aloof alsbery alter alterations altercation although altitude altman amass amateur amazement ambassador ambassadors ambition

Total words: **155,963** • Unique: **10,484** • Only in this book: **6,294**



Allen, James
Lane
*The Reign of
Law*

Bachelor,
Irving
D’Ri and I

Bachelor,
Irving
Eben Holden

Bova, Ben
*The Sam Gunn
Omnibus*

Brust, Steven
*My Own Kind
of Freedom*

Bryan,
Kathleen
*The Serpent
and the Rose*

Chestnutt,
Charles W
*The House
Behind the
Cedars*

Churchill,
Winston
Richard Carvel

de Lint,
Charles
Spiritwalk

Doctorow,
Cory
*Down and Out
in the Magic
Kingdom*

Doctorow,
Cory
*Eastern Stan-
dard Tribe*

Doct
Cory
*Som
Com
Town
one
Town*

Collins, Wilkie
The Moonstone

Moonstone. His offices were in the Duke Block, over the drug store. Larry, the Dr. Howard **Archie** had just come up from a game of pool with the Jewish clothier and two traveling men who happened to be staying overnight in

As the doctor in New England villages is proverbially old, so the doctor in small Colorado towns twenty-five years ago was generally young. Dr. **Archie** was barely thirty. He was tall, with massive shoulders which he held stiffly, and a

Denver tailor. The doctor was always well dressed.

Dr. **Archie** turned up the student’s lamp and sat down in the swivel chair before his desk. He sat uneasily, beating a tattoo on his knees with his fingers, and

smelling of warming flannels.

At three o’clock in the morning Dr. **Archie** was in the parlor putting on his cuffs and coat--there was no spare bedroom in that house. Peter **Kronborg’s**

tongue.

“Now, be still, I want to count.” Dr. **Archie** reached for her hand and took out his watch. When he put her hand back under the quilt he went over to one of the

When she heard the front door open, it occurred to her that the pleasant thing which was going to happen was Dr. **Archie** himself. He came in and warmed his hands at the stove. As he turned to her, she threw herself wearily toward him,

and so **milky** white. Thea must have got her hair and her silky skin from her mother. She was a little **Swede**, through and through. Dr. **Archie** could not help thinking how he would cherish a little creature like this if she were his. Her



CREATE A GROUP

Group name

e.g., "Floridian SciFi"

Authors

Books

Publication year

From to

Year of birth

From to

Year of death

From to

CREATE GROUP

COMPARE GROUPS

to

COMPARE

Sample Data

Author	Title	Publication year	Unique words	Words not found in Cory Doctorow's <i>Down and Out...</i>	% of words not shared
Allen, James Lane	<i>The Reign of Law</i>	1900	7,286	4,553	62
Bacheller, Irving	<i>D'Ri and I</i>	1901	6,159	3,630	59
Bacheller, Irving	<i>Eben Holden</i>	1900	7,585	4,768	63
Chestnutt, Charles W.	<i>The House Behind the Cedars</i>	1900	7,415	4,625	62
Churchill, Winston	<i>Richard Carvel</i>	1899	12,251	8,715	71
Dreiser, Theodore	<i>Sister Carrie</i>	1900	10,484	6,294	60
Ford, Paul Leicester	<i>Janice Meredith</i>	1901	11,907	8,491	71
Goss, Charles Frederic	<i>The Redemption of David Corson</i>	1900	9,099	5,989	66
Grant, Robert	<i>Unleavened Bread</i>	1900	9,418	6,346	67
Johnston, Mary	<i>Audrey</i>	1902	8,910	6,004	67
Johnston, Mary	<i>To Have and to Hold</i>	1900	8,130	5,346	66
Major, Charles	<i>Dorothy Vernon of Haddon Hall</i>	1902	7,294	4,562	63
Major, Charles	<i>When Knighthood Was in Flower</i>	1898	7,414	4,685	63
Norris, Frank	<i>The Octopus</i>	1901	12,790	9,014	70
Tarkington, Booth	<i>The Two Vanrevels</i>	1902	7,181	4,426	62
Thompson, Maurice	<i>Alice of Old Vincennes</i>	1900	9,664	6,393	66

Sample Data

Author	Title	Pub. year	Total word count	Unique words	% words unique	Words not in opposite set
Ringo, John	<i>A Hymn Before Battle</i>	2000	125,177	11,024	8.81	2,815
Lackey, Mercedes	<i>Beyond World's End</i>	2001	105,299	9,322	8.85	1,821
Link, Kelly	<i>Stranger Things Happen</i>	2001	81,494	6,837	8.39	1,220
Duberman, Martin	<i>Haymarket</i>	2003	112,885	11,480	10.17	1,949
Doctorow, Cory	<i>Down and Out in the Magic Kingdom</i>	2003	47,344	6,915	14.61	1,349
Doctorow, Cory	<i>Eastern Standard Tribe</i>	2004	50,133	7,284	14.53	1,732
Franco, Jorge	<i>Rosario Tijeras</i>	2004	44,700	3,932	8.80	403
Mamatas, Nick	<i>Move Under Ground</i>	2004	58,121	7,321	12.60	1,469
Watts, Peter	<i>Behemoth</i>	2004	152,976	13,392	8.75	4,149
Doctorow, Cory	<i>Someone Comes to Town, Someone Leaves Town</i>	2005	106,791	9,782	9.16	2,392
Allen, James Lane	<i>The Reign of Law</i>	1900	64,357	7,286	11.32	1,184
Bachelor, Irving	<i>D' Ri and I</i>	1901	65,047	6,159	9.47	1,067
Bachelor, Irving	<i>Eben Holden</i>	1900	88,115	7,585	8.61	1,419
Chestnutt, Charles W.	<i>The House Behind the Cedars</i>	1900	69,045	7,415	10.74	1,252
Churchill, Winston	<i>Richard Carvel</i>	1899	190,787	12,251	6.42	3,246
Ford, Paul Leicester	<i>Janice Meredith</i>	1901	182,739	11,907	6.52	3,219
Goss, Charles Frederic	<i>The Redemption of David Corson</i>	1900	100,234	9,099	9.08	1,681
Grant, Robert	<i>Unleavened Bread</i>	1900	122,573	9,418	7.68	1,918
Johnston, Mary	<i>Audrey</i>	1902	120,482	8,910	7.40	2,012
Johnston, Mary	<i>To Have and to Hold</i>	1900	115,406	8,130	7.04	1,752
Dreiser, Theodore	<i>Sister Carrie</i>	1900	155,963	10,484	6.72	1,771
Major, Charles	<i>Dorothy Vernon of Haddon Hall</i>	1902	116,277	7,294	6.27	1,214
Major, Charles	<i>When Knighthood Was in Flower</i>	1898	184,592	7,414	4.02	1,468
Norris, Frank	<i>The Octopus</i>	1902	195,700	12,790	6.54	2,937
Tarkington, Booth	<i>The Two Vanrevels</i>	1902	62,373	7,181	11.51	1,333
Thompson, Maurice	<i>Alice of Old Vincennes</i>	1900	105,442	9,664	9.17	1,940